An Improved Bayesian Permutation Entropy Estimator

Zachary Blanks ZDB6DZ@VIRGINIA.EDU
University of Virginia, USA

Donald E. Brown DEB@VIRGINIA.EDU
University of Virginia, USA

Marc A. Adams MARC.ADAMS@ASU.EDU
Arizona State University, USA

Siddhartha S. Angadi SSA2W@VIRGINIA.EDU
University of Virginia, USA

Abstract

We introduce a novel hierarchical Bayesian permutation entropy (PermEn) estimator designed to improve biomedical time series entropy assessments, especially for short signals. Unlike existing methods requiring a substantial number of observations or which impose restrictive priors, our non-centered, Wasserstein optimized hierarchical approach enables efficient MCMC inference and a broader range of PermEn priors. Evaluations on synthetic and secondary benchmark data demonstrate superior performance over the current state-of-the-art, including 13.33-63.67% lower estimation error, 8.16-47.77% lower posterior variance, and 47-60.83% lower prior construction error ($p \leq 2.42 \times 10^{-10}$). Applied to cardiopulmonary exercise test oxygen uptake signals, we reveal a previously unreported 1.55% (95% credible interval: [0.62%, 2.52%]) entropy difference between obese and lean subjects that diminishes as exercise capacity increases. For individuals capable of completing at least 7.5 minutes of testing, the 95% credible interval contained zero, suggesting potential insights into physiological complexity, exercise tolerance, and obesity. Our estimator refines biomedical signal PermEn estimation and underscores entropy's potential value as a health biomarker, opening avenues for further physiological and biomedical exploration.

Data and Code Availability We used three main datasets for this paper: a synthetically generated signal set, electrocardiogram (ECG) signals obtained from the 2001 paroxysmal atrial fibrillation (PAF) challenge dataset available in the PhysioNet database (Goldberger et al., 2000), and data obtained from a ramp-incremental treadmill cardiopulmonary exercise testing (CPET) protocol involving 603 participants (Age: 43.98 ± 9.46 years; BMI: 33.20 ± 6.85; Sex: 38.9% male, 61.1% female) (Adams et al., 2019). The CPET data is not publicly available.

The code used for benchmarking with the synthetic signal set and the ECG signal set are included as supplemental material, and the code for the CPET analysis is not available for public distribution.

Institutional Review Board (IRB) Per 45CFR46.104 the study was determined to be exempt by the University of Virginia IRB.

1. Introduction

Permutation entropy (PermEn) is a measure, first proposed by Bandt and Pompe (2002), for analyzing the predictability or irregularity of a time series signal. It has become essential in various biomedical applications, including but not limited to epilepsy monitoring (Veisi et al., 2007; Ra et al., 2021; Kbah et al., 2022), anaesthesia tracking (Olofsen et al., 2008; Jordan et al., 2008; Franka et al., 2023; Zanner et al., 2023), and heart rate variability and beat dynamics classification Bian et al. (2012); Yin et al. (2020). Formally, PermEn quantifies the probability of rank-ordered templates within a signal and computes the Shannon entropy of these permutations.

1.1. PermEn Computation Overview

PermEn computation is a two-step process. Initially, a time series signal $\boldsymbol{x} \in \mathbb{R}^N$, consisting of N observations, is transformed into a permutation symbol

count vector $\boldsymbol{c} \in \mathbb{Z}_+^{m!}$ with m referring to the "embedding dimension." This transformation effectively summarizes the original signal's patterns into a discrete set of symbols, aimed at capturing dynamical changes over time (Bandt and Pompe, 2002). Subsequently, PermEn is estimated by calculating the Shannon entropy of this symbol distribution.

The first stage of transforming a signal, $\boldsymbol{x}$, into a count vector, $\boldsymbol{c}$, is summarized in Algorithm 1.

Algorithm 1: Computing Permutation Symbol Counts

1. Instantiate $\boldsymbol{y} \in \Phi^{N-m+1}$ and $\boldsymbol{c} \in \mathbb{Z}_+^{m!}$
2. For $i = 1$ to $N - m + 1$
   (a) Template selection: $\boldsymbol{x}_m^{(i)} \coloneqq \boldsymbol{x}[i : (i+m-1)]$
   (b) $\widetilde{\boldsymbol{x}}_m^{(i)} \coloneqq$ Ascending rank ordering of $\boldsymbol{x}_m^{(i)}$
   (c) $y_i \coloneqq g\left(\widetilde{\boldsymbol{x}}_m^{(i)}\right)$, mapping $\widetilde{\boldsymbol{x}}_m^{(i)}$ to its permutation symbol
3. For $j = 1$ to $m!$
   (a) $c_j \coloneqq \sum_{i=1}^{N-m+1} \mathbb{1}\left[y_i = \phi_j\right]$
4. Return $\boldsymbol{c}$

In Algorithm 1, m determines the length of subvectors extracted from $\boldsymbol{x}$ and, consequently, the size of the permutation symbol set $\Phi = \{\phi_1, \ldots, \phi_{m!}\}$, representing all unique rank-ordered patterns or permutations. The cardinality of this set is $m!$ because there is factorial relationship between the length of a sequence and the number of ways it can be permuted. Each rank-ordered sub-vector is mapped to a unique permutation symbol through a function g. For a visual guide to this mapping and their connection to permutation symbols, refer to Figure 4 in Appendix A.

After computing $\boldsymbol{c}$, the most common way PermEn is estimated is via the MLE method for Shannon entropy. That is, for the j^{th} permutation symbol, ϕ_j, we estimate its probability of occurrence through the expression: $\widehat{\pi}_j = \frac{c_j}{N-m+1}$. This process is repeated for all $m!$ symbols and gives us the estimated permutation symbol probability vector $\widehat{\boldsymbol{\pi}}$. The PermEn, then follows from the definition of Shannon entropy:

$$\widehat{\mathcal{H}}_m^{\text{MLE}} = -\frac{1}{\log m!} \sum_{j=1}^{m!} \widehat{\pi}_j \log \widehat{\pi}_j. \tag{1}$$

In Equation (1), we bound $\widehat{\mathcal{H}}_m^{\text{MLE}} \in [0, 1]$ by rescaling by a factor of $1/\log m!$.

1.2. Challenges with Estimating PermEn in the "Under-Sampled" Regime

A crucial condition for accurate PermEn estimation using the MLE-based approach is that the number of observations, N, must be significantly greater than the number of possible permutation symbols, $m!$, to ensure accurate estimation of the probability of observing the permutation symbols given the count data (Bandt and Pompe, 2002; Riedl et al., 2013). However, the biomedical domain frequently encounters situations where $N \leq m!$, a scenario referred to as the "under-sampled" regime (Knudson and Pillow, 2013). This challenge becomes more pressing given the growing prevalence of short-duration physiological signals in modern medical practice due to the need for rapid diagnostics in clinical settings (Rahul and Sharma, 2022; Gupta et al., 2021; Rahul et al., 2021; Clifford et al., 2017). This is particularly salient when analyzing times series data from cardiopulmonary systems in patients or research participants where there can be significant breath-by-breath (or beat-to-beat) variability. These signals are typically averaged over longer time domains (15-second to minute-averages) to ease clinical analyses with attendant loss of information related to entropy that may be of value (Wasserman et al., 1987).

The information theory community has recognized the challenge of the under-sampled regime (and thus by proxy short-duration signals) and made various attempts to address this bias in entropy estimation using frequentist methods (Paninski, 2003; Hausser and Strimmer, 2009). However, these methods still suffer from the issue of negative bias, albeit to a lesser degree, do not allow one to incorporate prior knowledge about the system under investigation, and generally do not have a natural way of estimating uncertainty without additional computation or further parametric assumptions (Archer et al., 2014; Traversaro and Redelico, 2018; Ricci and Perinelli, 2022). To address these limitations, our paper introduces a novel Bayesian approach to PermEn estimation, particularly focusing on the under-sampled regime. Bayesian methods, in contrast to frequentist ones, allow for the

integration of prior information and provide a probabilistic framework to quantify uncertainty, and may be better suited to small-sample scenarios (McNeish, 2016).

Inspired by the work of Archer et al. (2014) and Nemenman et al. (2001), our Bayesian formulation offers two advantages over traditional frequentist methods: the ability to specify a prior over the PermEn space and the implicit incorporation of uncertainty bounds as a component of the estimation process. We believe this Bayesian formulation presents a meaningful advancement over traditional frequentist methods, particularly in addressing the challenges presented by the under-sampled regime in biomedical contexts.

In this study, we validated our proposed method using two main components: established signal set benchmarks and applying the model to CPET data. The benchmark data serves as a reference point for evaluating the estimation error and posterior variance in comparison to existing implementations. Meanwhile, the CPET data underscores the practical utility of our model.

Oxygen uptake ($\dot{V}O_2$) is the focus of our CPET investigation. Derived from the Fick equation, $\dot{V}O_2$ represents the product of cardiac output and whole-body oxygen extraction, offering a unified variable to examine multiple physiological systems. Notably, the importance of $\dot{V}O_{2,\text{Peak}}$ as an independent predictor of all-cause, cancer, and cardiovascular morbidity and mortality (Gaesser and Angadi, 2021), as well as the consequential data loss when excluding $\dot{V}O_2$ kinetics from analyses, motivated us to explore the entropy of $\dot{V}O_2$ signals aiming to study potential connections between obesity and biological complexity.

The contributions of our paper are threefold:

1. We introduce a novel Bayesian hierarchical model for PermEn, specifically designed for full Markov Chain Monte Carlo (MCMC) posterior approximation within standard embedding dimension sizes. Unlike existing Bayesian implementations, this model supports a wide spectrum of PermEn priors (see Sections 3.1 and 3.2).

2. Through synthetic and secondary data source benchmarking, we demonstrate the superior performance of our estimator in terms of estimation error, variance, and prior construction compared to existing methods (refer to Section 4.1).

3. By applying our method to CPET data, we contribute fresh insights into the relationship between obesity and biological complexity. This application showcases the real-world utility and relevance of our estimator, as elaborated in Section 4.2.

2. Bayesian PermEn Overview

Let $\boldsymbol{\pi}$ denote the probability vector for $m!$ permutation symbols, and $\boldsymbol{c}$, the permutation symbol count vector obtained from Algorithm 1. By Bayes' rule we have:

$$\underbrace{p(\boldsymbol{\pi} \mid \boldsymbol{c})}_{\text{Posterior}} \propto \underbrace{p(\boldsymbol{c} \mid \boldsymbol{\pi})}_{\text{Likelihood}} \times \underbrace{p(\boldsymbol{\pi})}_{\text{Prior}}. \tag{2}$$

We model the likelihood $p(\boldsymbol{c} \mid \boldsymbol{\pi})$ using a multinomial distribution: $\boldsymbol{c} \mid \boldsymbol{\pi} \sim \text{Mult}(N - m + 1, \boldsymbol{\pi})$, and approximate the PermEn distribution $p(\mathcal{H}_m, \mid \boldsymbol{c})$ using the posterior, $p(\boldsymbol{\pi} \mid \boldsymbol{c})$. Given a specific probability vector, PermEn is deterministic.

Unfortunately, analytic derivation of $p(\boldsymbol{\pi} \mid \boldsymbol{c})$ is typically infeasible. Nevertheless, sampling from $p(\boldsymbol{\pi} \mid \boldsymbol{c})$ enables the approximation of $p(\mathcal{H}_m \mid \boldsymbol{c})$. Thus, the critical step in Bayesian PermEn estimation involves defining a suitable prior, $p(\boldsymbol{\pi})$, considering two primary challenges.

First, the high-dimensionality of $\boldsymbol{\pi}$ complicates the conceptualization of a "reasonable" prior, as it is impractical to favor specific permutation symbols a priori. Second, and more crucially, is the consideration of what $p(\boldsymbol{\pi})$ implies for the PermEn prior, $p(\mathcal{H}_m)$, given the nonlinear and non-invertible relationship between probability vectors and Shannon entropy.

2.1. Bayesian PermEn Conjugate Prior Models

A common specification of $p(\boldsymbol{\pi})$ is to leverage Bayesian conjugate priors. Notably, Little et al. (2022) and Pose et al. (2021) advocate for the prior: $\boldsymbol{\pi} \sim \text{Dir}(\boldsymbol{\alpha})$ with $\boldsymbol{\alpha} \in \mathbb{R}^{m!}_{++}$ being a strictly positive concentration parameter. This choice allows for the exploitation of the Dirichlet-Multinomial conjugacy, yielding a posterior distribution $\boldsymbol{\pi} \mid \boldsymbol{c} \sim \text{Dir}(\boldsymbol{\alpha} + \boldsymbol{c})$. Given the posterior's analytical form, sampling to approximate $p(\mathcal{H}_m \mid \boldsymbol{c})$ is straightforward.

Nevertheless, a critical aspect of this model is selecting the concentration parameter, $\boldsymbol{\alpha}$. Common choices are the Laplace prior ($\alpha = 1$; repeated for all $m!$ components of the vector) and Perks' prior ($\alpha = \frac{1}{m!}$; similarly repeated for all $m!$ components).

The former implies a uniform prior over the probability simplex, while the latter approximates an objective Bayesian prior, as discussed by Berger et al. (2015) and thus has tended to be favored in the literature (Pose et al., 2021). However, Nemenman et al. (2001) demonstrated that fixed values for $\boldsymbol{\alpha}$ result in highly informative implied entropy priors that become increasingly concentrated in higher dimensions.

2.2. NSB PermEn Estimator

To address the limitations of fixed Dirichlet priors, Nemenman et al. (2001) proposed a non-informative, hierarchical prior for the Dirichlet concentration parameter, α:

$$p(\alpha) \propto \frac{d}{d\alpha}\mathbb{E}\left[\mathcal{H}_m \mid \alpha\right] = m! \cdot \psi_1(m! \cdot \alpha + 1) - \psi_1(\alpha + 1) \tag{3}$$

where $\psi_1(\cdot)$ is the tri-gamma function. They demonstrated that Equation (3) implied an approximately uniform prior for $\mathcal{H}_m \in [0, \log m!]$ which can be rescaled to $[0, 1]$ to align with the PermEn set-up without affecting the distribution.

The so-called NSB entropy estimator applied to the PermEn is then given by:

$$\widehat{\mathcal{H}}_m^{\text{NSB}} = \frac{1}{\log m!} \int \mathbb{E}\left[\mathcal{H}_m \mid \boldsymbol{y}, \alpha\right] \frac{p(\boldsymbol{y} \mid \alpha) p(\alpha)}{p(\boldsymbol{y})}\, d\alpha \tag{4}$$

Archer et al. (2014) provide analytic expressions for $\mathbb{E}\left[\mathcal{H}_m \mid \boldsymbol{y}, \alpha\right]$ and $\mathbb{E}\left[\mathcal{H}_m^2 \mid \boldsymbol{y}, \alpha\right]$, which can be used with one-dimensional numerical integration to efficiently compute the first and second moment of $p\left(\mathcal{H}_m \mid \boldsymbol{c}\right)$.

The NSB entropy estimator, while advantageous in the under-sampled regime and noted for its favorable bias and variance properties (Nemenman et al., 2001; Archer et al., 2014), has certain limitations.

2.2.1. NSB Estimator Limitations

The NSB estimator faces two main challenges: its dependence on a non-informative prior for entropy, and the computational demands of estimating the complete entropy posterior distribution.

In many scenarios, especially in the biomedical signal domain, researchers may have valuable insights into the expected range of PermEn values. For example, in prior research analyzing CPET signals, Blanks et al. (2024) observed that certain signals, such as $\dot{V}O_2$, exhibited higher entropy compared to well-known benchmarks like sinusoidal signals. In such cases, it becomes beneficial to incorporate this knowledge into the PermEn inference process. Furthermore, in the context of studying phenomena with limited sample sizes, the idea of a non-informative prior can be more of a theoretical concept than a practical reality (Gelman, 2006). As McNeish (2016) states, "...when the information contained in the likelihood is relatively small due to a limited sample size, the prior will necessarily play a key role in the posterior distribution".

Second, although Nemenman et al. (2001) and Archer et al. (2014) offer a method for estimating the posterior PermEn distribution's first and second moments, full posterior approximation via MCMC remains cumbersome. The absence of an efficient sampling algorithm for the NSB prior, coupled with potential numerical instabilities from estimating gradients of the tri-gamma function, leads to less efficient, gradient-free sampling techniques like Metropolis-Hastings. See Appendix D for more details.

3. Improving the Bayesian PermEn Model

We introduce a new Bayesian estimator for PermEn. Our method not only accommodates a broader class of PermEn priors, allowing for the integration of domain-specific knowledge, but also enhances computational efficiency in approximating the complete posterior PermEn distribution.

3.1. An Alternate Hierarchical Bayesian PermEn Prior

We desire a prior that is both numerically stable and amenable to Hamiltonian Monte Carlo sampling methods. We propose to address these challenges by employing the hierarchical prior:

$$\begin{aligned} \alpha &\sim \Gamma(a, b) \\ \boldsymbol{\epsilon} &\sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{I}_{m!}\right) \\ \widetilde{\boldsymbol{\pi}} \mid \alpha, \boldsymbol{\epsilon} &\coloneqq \alpha \cdot \boldsymbol{\epsilon} \\ \boldsymbol{\pi} \mid \widetilde{\boldsymbol{\pi}} &\coloneqq \operatorname{softmax}\left(\widetilde{\boldsymbol{\pi}}\right) \end{aligned} \tag{5}$$

There are three key advantages to this prior construction compared to existing implementations. First, using a Gamma distribution for the concentration hyperprior, $\alpha \sim \Gamma(a, b)$, allows us to employ

numerically efficient sampling algorithms versus the kernel specified in Equation (3). Second, sampling in the logit space via: $\widetilde{\boldsymbol{\pi}} \coloneqq \alpha \cdot \boldsymbol{\epsilon}$ and then transforming the logit into a valid probability vector using the softmax function circumvents issues associated with unfavorable posterior geometry by exploring near the probability simplex boundary. Third, using a non-centered hierarchical model versus a centered formulation: $\widetilde{\boldsymbol{\pi}} \sim \mathcal{N}(\mathbf{0}, \alpha \cdot \boldsymbol{I}_{m!})$, may aid posterior approximation in limited data regimes (Papaspiliopoulos et al., 2007).

3.2. Optimizing the PermEn Hierarchical Prior

Our goal is to ensure the implied entropy prior $p(\mathcal{H}_m)$ from our hierarchical model closely matches a specified reference prior distribution η. We achieve this by manipulating the Gamma hyperprior parameters (a, b) and measure closeness using the Wasserstein-1 (W_1) distance between $p(\mathcal{H}_m)$ and η.

The W_1 distance can be approximated numerically from samples as:

$$W_1(\mathcal{H}_m, \eta) \approx \sum_{i=2}^{D+L} |\widehat{F}_{\mathcal{H}_m}(z_i) - \widehat{F}_{\eta}(z_i)| \Delta z_i, \tag{6}$$

where $\widehat{F}_{\mathcal{H}_m}$ and $\widehat{F}_{\eta}$ are the empirical cumulative distribution functions of $\mathcal{H}_m$ and η, respectively. The vector $\boldsymbol{z}$ is the merged, sorted set of all samples from both distributions. Assuming the number of draws are sufficiently large to represent their respective probability measures, then Equation (6) is a good numerical approximation of the W_1 distance (Peyré and Cuturi, 2019).

We optimize the Gamma parameters to minimize this W_1 approximation, using Algorithm 2. Further optimization details are provided in Appendix B with example results depicted graphically in Figure 1.

3.3. Markov Chain Monte Carlo Approximation of Posterior PermEn

Given optimal hyperprior parameters (a^*, b^*) for a reference measure, η, under our proposed hierarhical prior, the posterior is:

$$\begin{aligned} p(\boldsymbol{\pi} \mid \boldsymbol{c}) &\propto p(\boldsymbol{c} \mid \boldsymbol{\pi})\, p(\boldsymbol{\pi}) \\ &\propto \prod_{j=1}^{m!} \pi_j^{c_j} \iint p(\boldsymbol{\pi} \mid \alpha, \boldsymbol{\epsilon}, a^*, b^*)\, d\alpha\, d\boldsymbol{\epsilon}. \end{aligned} \tag{7}$$

Algorithm 2: PermEn Hyperprior Parameter Optimization

1. For $t = 1$ to T gradient descent steps
   (a) $\left\{\alpha^{(d)}\right\}_{d=1}^{D} \sim \text{Gamma}(a^{(t)}, b^{(t)})$
   (b) $\left\{\boldsymbol{\epsilon}^{(d)}\right\}_{d=1}^{D} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{m!})$
   (c) $\left\{\boldsymbol{\pi}^{(d)}\right\} \coloneqq \text{softmax}\left(\alpha^{(d)} \cdot \boldsymbol{\epsilon}^{(d)}\right) \quad \forall d$
   (d) $\left\{h_{\text{Prior}}^{(d)}\right\} \coloneqq -\frac{1}{\log m!} \sum_{j=1}^{m!} \pi_j^{(d)} \log \pi_j^{(d)} \quad \forall d$
   (e) $\boldsymbol{z} \coloneqq \text{mergesort}\left(\left\{\eta^{(d)}\right\}_{d=1}^{L}, \left\{h_{\text{Prior}}^{(d)}\right\}_{d=1}^{D}\right)$
   (f) $\Delta \boldsymbol{z} \coloneqq (z_{k+1} - z_k)_{k=2}^{D+L}$
   (g) $W_1(\mathcal{H}_m, \eta) \approx \sum_{i=2}^{D+L} |\widehat{F}_{H_m}(z_i) - \widehat{F}_{\eta}(z_i)| \Delta z_i$
   (h) Backpropagation update:
      i. $a^{(t+1)} \coloneqq a^{(t)} - \lambda \frac{\partial W_1}{\partial a}$
      ii. $b^{(t+1)} \coloneqq b^{(t)} - \lambda \frac{\partial W_1}{\partial b}$
2. Return optimal Gamma parameters: (a^*, b^*)

Direct analytical computation of this posterior is infeasible due to the integrals over $m! + 1$ dimensions. We address this through Markov Chain Monte Carlo (MCMC) sampling, which bypasses the need for explicit integral solutions by generating posterior distribution samples. This method enables estimation of means, variances, and credible intervals (CIs), effectively converting the task of solving complex integrals into sample set summarization (Betancourt, 2018). Algorithm 3 details the process for estimating the posterior PermEn distribution and then calculating quantities of interest.

Using the PyMC framework (Abril-Pla et al., 2023), we validated MCMC convergence via trace plots, effective sample size checks, and the Gelman-Rubin "R-hat" statistic (Gelman and Rubin, 1992).

4. Results

In this section, we present the results obtained from a series of synthetic and real-world experiments conducted to evaluate the performance of our proposed Bayesian PermEn estimator. Our primary focus is on assessing its performance in the under-sampled regime.

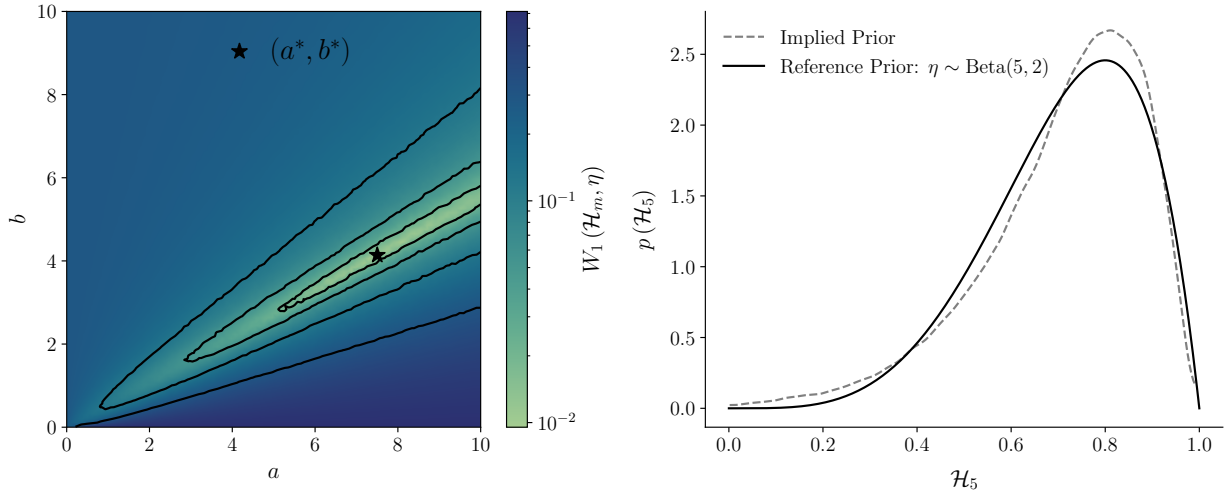

Figure 1: W_1 distance loss surface under $\eta \sim \text{Beta}(5, 2)$ reference prior at $m = 5$ (left), and the resulting implied PermEn prior (right)

Algorithm 3: Posterior PermEn MCMC Approximation

1. Compute the permutation symbol count vector $\boldsymbol{c}$ from time series data $\boldsymbol{x}$ using Algorithm 1
2. Compute (a^*, b^*) using Algorithm 2 for a reference prior η
3. Use the No U-Turn MCMC Sampler (Homan and Gelman, 2014) to obtain posterior samples: $\left\{\boldsymbol{\pi}^{(d)}\right\}_{d=1}^{D} \sim p(\boldsymbol{\pi} \mid \boldsymbol{c})$ under Equation (7)
4. Calculate posterior PermEn: $\left\{h^{(d)}\right\} := -\frac{1}{\log m!} \sum_{j=1}^{m!} \pi_j^{(d)} \log \pi_j^{(d)} \quad \forall d$
5. $\widehat{\mathcal{H}}_m^{\text{Bayes}} = \mathbb{E}\left[\mathcal{H}_m \mid \boldsymbol{c}\right] \approx \frac{1}{D} \sum_{i=d}^{D} h^{(d)}$
6. $\widehat{\sigma}_m^{\text{Bayes}} = \sqrt{\frac{1}{D} \sum_{d=1}^{D} \left(h^{(d)} - \widehat{\mathcal{H}}_m^{\text{Bayes}}\right)^2}$
7. Estimate the γ% CI using $\left\{h^{(i)}\right\}_{d=1}^{D}$ (Chen and Shao, 1999)
8. Return $\widehat{\mathcal{H}}_m^{\text{Bayes}}$, $\widehat{\sigma}_m^{\text{Bayes}}$, and the γ% CI

4.1. Benchmark Signals Experiments

We evaluated our proposed Bayesian PermEn estimator's performance relative to existing methods including the NSB estimator and a conjugate Bayesian model under Perks' prior. This analysis used synthetic autoregressive AR(1) signals and real-world ECG signals.

While the true PermEn cannot be calculated exactly for finite signals, we obtained sufficient approximations for both signal classes.

For AR(1) processes, we generated a signal of length $N = 1,000,000$ observations, and for the ECG signals, we considered a five-minute segment sampled at 128 Hz, resulting in a signal with $N = 38,400$ observations. For embedding dimensions $m = 5$ and $m = 6$, these lengths exceed the heuristics $N > 5m!$ and $N > 10m!$ where maximum likelihood PermEn estimates stabilize Amigó et al. (2008); Cuesta-Frau et al. (2019). Hence, we treated the maximum likelihood estimates as adequate "true" PermEn approximations.

We randomly selected 50 sub-signals of varying lengths $N = \{10, 20, 30, 40, 50\}$ from each signal class and estimated PermEn using:

- Our estimator with a non-informative reference prior, $\eta \sim \text{Beta}(1, 1)$, matching NSB's implied prior.

- Our estimator with an informative reference prior: $\eta \sim \text{Beta}(10, 2)$.
- The NSB estimator
- A conjugate Bayesian estimator with Perks' prior ($\boldsymbol{\pi} \sim \text{Dir}\left(\frac{1}{m!}\right)$).

Let $\widehat{\mathcal{H}}_m^{(i,N,e)}$ be the i-th expected posterior PermEn estimate and $\widehat{\sigma}_m^{(i,N,e)}$ is the posterior standard deviation obtained from a sub-signal containing N observations using Bayesian PermEn estimator e at embedding dimension m.

We evaluated the root mean squared error (RMSE) by calculating:

$$\text{RMSE}(N, e, m) = \sqrt{\frac{1}{50}\sum_{i=1}^{50}\left(\widehat{\mathcal{H}}_m^{(i,N,e)} - \mathcal{H}_m^{\text{True}}\right)^2},$$

where $\mathcal{H}_m^{\text{True}}$ represents the "true" PermEn, derived from the complete signal. Similarly, the mean posterior uncertainty is given by:

$$\overline{\sigma}(N, e, m) = \frac{1}{50}\sum_{i=1}^{50}\widehat{\sigma}_m^{(i,N,e)}.$$

Figure 2 shows the results.

There are a number of takeaways from this experiment. First, as expected, increasing the signal length N resulted in decreased RMSE and mean posterior uncertainty across all estimators, which aligns with the expected behavior of Bayesian estimators.

Second, the conjugate Bayesian estimator with Perks' prior consistently performed poorest across signal types, embedding dimensions, and N. This stems from Perks' narrow implied PermEn prior range. When the true PermEn falls outside this range, as in our analysis, estimator accuracy suffers. Furthermore, the Perks' prior estimator has lower variance when transitioning from $m = 5$ to $m = 6$, indicating that not only is the estimator inaccurate, but it is also confidently so. In contrast, while transitioning from $m = 5$ to $m = 6$ greatly increases dimensionality (120 to 720 dimensions), the mean posterior variance increase was relatively modest for our estimator and NSB.

Intriguingly, our estimator under a $\eta \sim \text{Beta}(1, 1)$ reference prior – "Logit(1, 1)" – outperformed NSB for most cases despite both implying non-informative priors, achieving 9.52-20.00% (AR(1)) and 2.22-11.39% (PAF) lower RMSE. We attribute this to numerical differences in the reference priors, particularly in the tails. Figure 5 shows NSB overestimating the $p(\mathcal{H}_5) = 1$ density by more than a factor of two, likely causing its posterior to favor lower values unduly. Appendix C provides further details on this discrepancy between the two ostensibly equivalent non-informative priors.

Fourth, our estimator with an $\eta \sim \text{Beta}(10, 2)$ reference prior – Logit(10, 2) – consistently achieved the lowest RMSE and mean posterior variance compared to Logit(1, 1) estimator and NSB for both signal classes. For AR(1) signals, it had 28.57 to 63.67% lower RMSE and 21.05 to 57.92 lower variance than NSB, and 9.52-20% lower RMSE than Logit(1, 1) with similar variance improvements. For the PAF signals, it had 13.33-51.52% lower RMSE and 8.16-39.18% lower variance than NSB, and 8.86-46.64% lower RMSE than the Logit(1, 1) model. Incorporating an informative prior guided inference more effectively, reducing error and variance. This highlights a key advantage of our estimator's versatility in capturing diverse priors beyond just non-informative ones.

4.2. CPET Analysis

Cardiopulmonary exercise testing (CPET) is a non-invasive procedure used to measure the integrated response of the heart, lungs, and muscles to graded exercise. It provides comprehensive physiological measurements in response to exercise, on a breath-by-breath basis, including cardiovascular variables (e.g., heart rate, blood pressure), ventilatory variables (breathing frequency, tidal volume, minute ventilation) and gas exchange variables such as oxygen uptake($\dot{\text{V}}\text{O}_2$) and carbon dioxide output ($\dot{\text{V}}\text{CO}_2$) (Wasserman et al., 1987). $\dot{\text{V}}\text{O}_2$ during CPET is the volume of oxygen taken up at the mouth per unit time, that is subsequently used for aerobic metabolism to support the increased energy demands of the task. Greater $\dot{\text{V}}\text{O}_2$ indicates more capacity to perform physical work (Bassett and Howley, 2000).

We analyzed $\dot{\text{V}}\text{O}_2$ signals from ramp-incremental treadmill CPETs to answer:

- Are there PermEn differences between obese and lean participants?
- If there is a difference, does this difference change the longer a participant is able to endure a CPET?

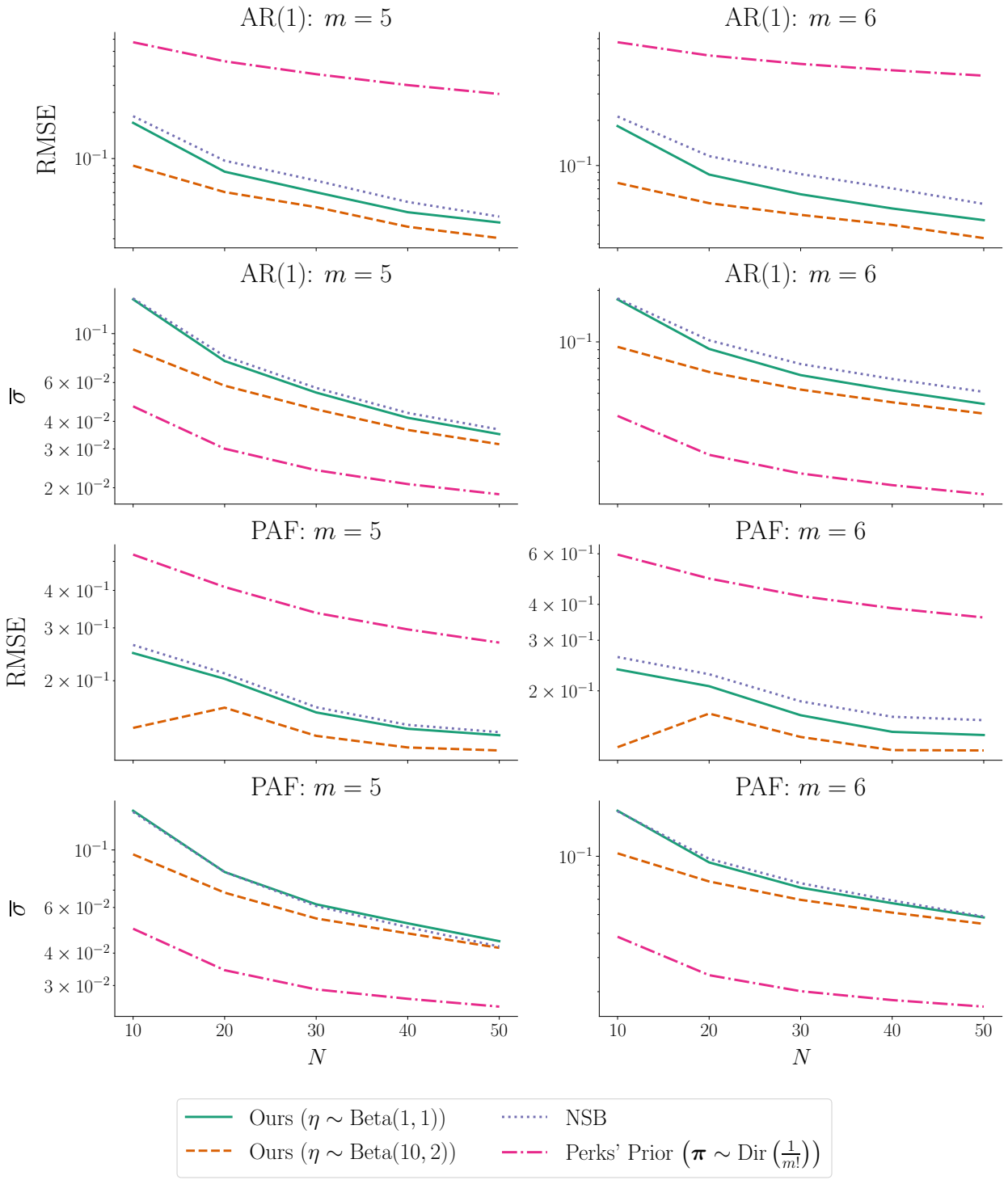

Figure 2: Our proposed Bayesian PermEn estimator consistently outperformed competing approaches in terms of RMSE and had tighter posterior variance compared to the NSB model.

We ensured CPET signal statistical stationarity at $\alpha = 0.05$, a crucial condition for valid entropy analysis (Chatain et al., 2020). To achieve weakly stationary signals, we implemented Gaussian process detrending yielding zero-mean, unit-variance signals. Subsequently, we applied the Augmented Dickey-Fuller test with Holm-Sidak correction for multiple comparisons and identified 317 weakly stationary $\dot{V}O_2$ signals (Dickey and Fuller, 1979; Holm, 1979; Šidák, 1967).

The remaining signals had a median length of $N = 40$ observations, 78% had $N \leq 50$ observations, all signals contained fewer than 100 samples. We estimated PermEn at $m = 6$ using a Beta(5,2) prior balancing estimate uncertainty stemming from short signal lengths with previous entropy estimates (Blanks et al., 2024; Riedl et al., 2013). Participants with a body mass index greater than 30 were classified as obese.

For each obesity status $o \in \{\text{Lean}, \text{Obese}\}$, let $\mathcal{H}(o) = \left\{\left(\widehat{\mathcal{H}}_m^{(i,o)}, \widehat{\sigma}^{(i,o)}\right)\right\}_{i=1}^{n_o}$ be the posterior PermEn estimates and variances. We compared PermEn of $\dot{V}O_2$ between obese and lean participants using a Bayesian model which treats the estimates as noisy measurements with known variance (Figure 3).

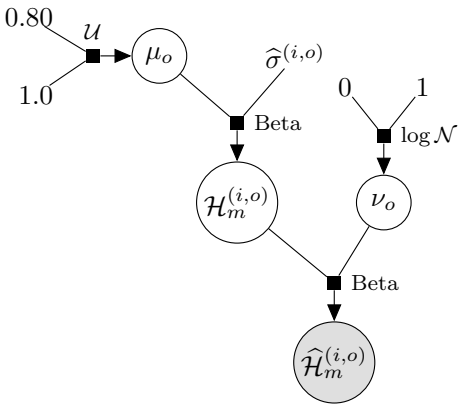

Figure 3: Bayesian model for comparing the PermEn of obese versus lean participants.

Let the mean posterior PermEn be μ_0 for lean individuals and μ_1 for obese participants. The relative percentage decrease in entropy between groups is:

$$\Delta\mu = \left(\frac{\mu_0 - \mu_1}{\mu_0}\right) \times 100.$$

Positive $\Delta\mu$ indicates, on average, greater relative entropy for lean individuals.

We inferred μ_0, μ_1, and $\Delta\mu$ for participants able to exercise at least 2.5, 5, 7.5, and 10 minutes, corresponding to $\dot{V}O_2$ signals with at least 10, 20, 30, and 40 observations, respectively. We hypothesized lean individuals would have higher relative entropy that diminishes with prolonged exercise duration. Results are in Table 1.

The analysis supports both hypotheses. First, using the complete signal set ($N \geq 10$ observations, $\geq$ 2.5 minute CPET), lean individuals exhibit approximately 1.55% higher relative PermEn than obese participants on average. Furthermore, as exercise duration increases, the relative entropy difference diminishes by 9.2%, 43.3%, and 53.8% for particpants who completed at least 5, 7.5, and 10 minute CPETs, respectively. Beyond 7.5 minutes, the 95% credible interval for $\Delta\mu$ contains zero.

This finding may have clinical implications, suggesting that the relative effect of obesity diminishes as an individual's exercise capacity increases.

5. Discusssion

In this study, we evaluated our Bayesian PermEn estimator in comparison to existing implementations, utilizing benchmark signals and real-world CPET data.

In the benchmark signal analysis, we assessed our estimator's performance across various embedding dimensions and signal lengths. We calculated the RMSE and mean posterior variance, revealing several key insights. Our approach consistently outperformed others, yielding the lowest RMSE. Notably, when we provided a more informative PermEn prior, we achieved both reduced error and smaller variance, underscoring the advantages of informative priors in small-sample scenarios. Additionally, we observed that the increase in embedding dimension from $m = 5$ to $m = 6$ resulted in a slight increase in posterior variance for both our approach and the NSB estimator. However, this was not the case for the conjugate Bayesian model with Perks' prior, a concerning outcome given the substantial increase in dimensionality.

Furthermore, despite the similarity in the implied PermEn prior between our model under a reference prior of $\eta \sim \text{Beta}(1, 1)$ and the NSB estimator, our model consistently achieved lower RMSE and comparable posterior variance. This outcome appeared to be driven by reduced target overshooting near the

Table 1: Lean individuals exhibit higher relative PermEn than obese individuals that diminishes with increasing exercise duration.

Minimum CPET Length (min)	Remaining Participants	μ_0 Mean	μ_0 95% CI	μ_1 Mean	μ_1 95% CI	$\Delta\mu$ Mean	$\Delta\mu$ 95% CI
2.5	317	0.914	[0.907, 0.920]	0.899	[0.894, 0.905]	1.551%	[0.618%, 2.518%]
5	307	0.914	[0.907, 0.920]	0.901	[0.895, 0.906]	1.409%	[0.462%, 2.339%]
7.5	260	0.914	[0.908, 0.921]	0.906	[0.900, 0.912]	0.880%	[−0.057%, 1.843%]
10	197	0.917	[0.911, 0.924]	0.911	[0.904, 0.918]	0.716%	[−0.333%, 1.804%]

lower and upper boundaries of the $[0, 1]$ interval. This discrepancy in implied priors may also explain why the NSB estimator tended to have slightly higher error, as it disproportionately favored lower entropy relative to the desired reference uniform PermEn prior.

Next, we applied our Bayesian PermEn estimator to real-world CPET data, comparing the entropy of $\dot{V}O_2$ between lean and obese individuals. Across the entire signal set, lean individuals had approximately 1.55% greater relative entropy compared to obese participants. However, by increasing the required minimum signal length (assessing longer minimum CPET lengths), effectively excluding the unhealthiest individuals, we observed that the relative difference in entropy between lean and obese individuals diminished. If an individual could endure a CPET for at least 7.5 minutes, the 95% credible interval contained zero.

Our study's results highlight two key advantages of our Bayesian PermEn estimator compared to the NSB estimator and conjugate Bayesian models with fixed Dirichlet concentration priors. One, our approach allows for a broader class of PermEn priors over $[0, 1]$, and our computational optimization method constructs numerically superior priors relative to existing implementations. Two, by employing a non-centered and transformed PermEn formulation, we efficiently use MCMC to sample from a parameter concentration hyperprior, enabling the approximation of the complete PermEn posterior distribution rather than being limited to its first and second moments.

5.1. Clinical Implications

CPET plays a vital role in evaluating cardiopulmonary function for a range of clinical applications including assessment of undiagnosed exercise intolerance, heart failure, unexplained dyspnea, monitoring therapy in chronic lung disease and heart failure and for staging of cardiac transplantation (Balady et al., 2010). However, CPET signals, characterized by increasing work rate, are inherently nonstationary which poses challenges for nonlinear dynamical analysis techniques like PermEn that assume stationarity (Chatain et al., 2020).

Our Bayesian PermEn framework enables reliable entropy estimation even for the relatively short CPET signals containing fewer than 50 observations that are commonly encountered clinically after accounting for stationarity. This approach could expand investigational research into the complex physiological mechanisms governing CPET responses across different age groups and diseases. For instance, previous studies have suggested differences in heart rate complexity between healthy individuals and those with congestive heart failure (Zhao et al., 2015). Our methodology helps mitigate the bias and uncertainty stemming from limited CPET signals to better understand if true physiological differences underlie these entropy differences. Additionally, improved precision of entropy estimation may provide greater insight into quantifying changes in neuromuscular fatigue development and skeletal muscle efficiency with aging and disease across integrated CPET parameters (Grassi et al., 2015).

Furthermore, our finding that the entropy gap between obese and lean individuals diminishes for those capable of continuing the CPET for over 7.5 minutes warrants further investigation. If confirmed in larger studies, this could suggest that encouraging obese patients to improve fitness enough to sustain longer CPET may impart beneficial effects. For example, large systematic reviews and meta-analyses have demonstrated that obese individuals with high $\dot{V}O_{2,\text{Peak}}$ (analogous to longer CPET duration) have

half the mortality risk of lean individuals with low $\dot{V}O_{2,Peak}$ and similar mortality outcomes compared to lean individuals with high $\dot{V}O_{2,Peak}$ (Barry et al., 2018).

Overall, the flexibility of our Bayesian PermEn technique coupled with its reliable uncertainty quantification opens promising avenues for elucidating the clinical implications of exercise stress testing across diverse populations and potentially for the development of a novel biomarker.

Acknowledgments

This work was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award UL1TR003015, and by the National Cancer Institute of the National Institutes of Health under Grant R01CA198915.

References

Oriol Abril-Pla, Virgile Andreani, Colin Carroll, Larry Dong, Christopher J. Fonnesbeck, Maxim Kochurov, Ravin Kumar, Junpeng Lao, Junpeng Lao, Osvaldo A. Martin, Michael Osthege, Ricardo Vieira, Thomas Wiecki, and Robert Zinkov. Pymc: a modern, and comprehensive probabilistic programming framework in python. *PeerJ Comput Science*, 9, 9 2023. doi: http://dx.doi.org/10.7717/peerj-cs.1516.

Marc A. Adams, Jane C. Hurley, Christine B. Phillips, Michael Todd, Siddhartha S. Angadi, Vincent Berardi, Melbourne F. Hovell, and Steven Hooker. Rationale, design, and baseline characteristics of walkit arizona: A factorial randomized trial testing adaptive goals and financial reinforcement to increase walking across higher and lower walkable neighborhoods. *Contemporary Clinical Trials*, 81:87–101, 6 2019. doi: https://doi.org/10.1016/j.cct.2019.05.001.

José María Amigó, Samuel Zambrano, and Miguel AF Sanjuán. Combinatorial detection of determinism in noisy time series. *Europhysics Letters*, 83(6):60005, 9 2008. doi: https://doi.org/10.1209/0295-5075/83/60005.

Evan Archer, Il Memming Park, and Jonathan W. Pillow. Bayesian entropy estimation for countable discrete distributions. *The Journal of Machine Learning Research*, 15(1):2833–2868, 1 2014.

Gary J. Balady, Ross Arena, Kathy Sietsema, Jonathan Myers, Lola Coke, Gerald F. Fletcher, Daniel Forman, Barry Franklin, Marco Guazzi, Martha Gulati, Steven J. Keteyian, Carl J. Lavie, Richard Macko, Donna Mancini, Council on Epidemiology Richard V. Milani, Council on Peripheral Vascular Disease Prevention, on behalf of Interdisciplinary Council on Quality of Care, and Outcomes Research. Clinician's guide to cardiopulmonary exercise testing in adults: A scientific statement from the american heart association. *Circulation*, 122(2):191–225, 7 2010. doi: https://doi.org/10.1161/CIR.0b013e3181e52e69.

Christoph Bandt and Bernd Pompe. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.*, 88(17):174102, 4 2002. doi: https://doi.org/10.1103/PhysRevLett.88.174102.

Vaughn W. Barry, Jennifer L. Caputo, and Minsoo Kang. The joint association of fitness and fatness on cardiovascular disease mortality: A meta-analysis. *Progress in Cardiovascular Diseases*, 61 (2):136–141, 7 2018. doi: https://doi.org/10.1016/j.pcad.2018.07.004.

David R Bassett and Edward T Howley. Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Medicine and Science in Sports and Exercise*, 32(1):70–84, 1 2000. doi: https://doi.org/10.1097/00005768-200001000-00012.

Alexandre Belloni and Victor Chernozhukov. On the computational complexity of mcmc-based estimators in large samples. *The Annals of Statistics*, 37 (4):2011–2055, 2009.

James O. Berger, Jose M. Bernardo, and Dongchu Sun. Overall objective priors. *Bayesian Analysis*, 10(1):189 – 221, 3 2015. doi: https://doi.org/10.1214/14-BA915.

Michael Betancourt. A conceptual introduction to hamiltonian monte carlo. *arXiv e-Prints*, 7 2018. doi: https://doi.org/10.48550/arXiv.1701.02434.

Chunhua Bian, Chang Qin, Qianli D. Y. Ma, and Qinghong Shen. Modified permutation-entropy analysis of heartbeat dynamics. *Phys. Rev. E*, 85:021906, 2 2012. doi: https://doi.org/10.1103/PhysRevE.85.021906.

Zachary Blanks, Donald E Brown, Dan M Cooper, Shlomit Radom Aizik, and Ronen Bar-Yoseph. Signal variability comparative analysis of healthy early- and late-pubertal children during cardiopulmonary exercise testing. *Medicine and Science in Sports and Exercise*, 56(2):287–296, 2 2024. doi: https://doi.org/10.1249/mss.0000000000003296.

Cyril Chatain, Mathieu Gruet, Jean-Marc Vallier, and Sofiane Ramdani. Effects of nonstationarity on muscle force signals regularity during a fatiguing motor task. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(1):228–237, 2020. doi: https://doi.org/10.1109/TNSRE.2019.2955808.

Ming-Hui Chen and Qi-Man Shao. Monte carlo estimation of bayesian credible and hpd intervals. *Journal of Computational and Graphical Statistics*, 8(1):69–92, 3 1999. doi: https://doi.org/10.2307/1390921.

Gari D Clifford, Chengyu Liu, Benjamin Moody, Li-wei H. Lehman, Ikaro Silva, Qiao Li, A E Johnson, and Roger G. Mark. Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017. In *2017 Computing in Cardiology (CinC)*, pages 1–4, 2017. doi: https://doi.org/10.22489/CinC.2017.065-469.

David Cuesta-Frau, Juan Pablo Murillo-Escobar, Diana Alexandra Orrego, and Edilson Delgado-Trejos. Embedded dimension and time series length. practical influence on permutation entropy and its applications. *Entropy*, 21(4), 4 2019. doi: https://doi.org/10.3390/e21040385.

David A. Dickey and Wayne A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431, 6 1979. doi: https://doi.org/10.2307/2286348.

Michelle Franka, Alexander Edthofer, Andreas Körner, Sandra Widmann, Thomas Fenzl, Gerhard Schneider, and Matthias Kreuzer. An in-depth analysis of parameter settings and probability distributions of specific ordinal patterns in the shannon permutation entropy during different states of consciousness in humans. *Journal of Clinical Monitoring and Computing*, 38(2):385–397, 7 2023. doi: https://doi.org/10.1007/s10877-023-01051-z.

Glenn A Gaesser and Siddhartha S Angadi. Obesity treatment: Weight loss versus increasing fitness and physical activity for reducing health risks. *iScience*, 24(10):102995, 10 2021. doi: https://doi.org/10.1016%2Fj.isci.2021.102995.

Andrew Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515 – 534, 9 2006. doi: https://doi.org/10.1214/06-BA117A.

Andrew Gelman and Donald B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457 – 472, 11 1992. doi: https://doi.org/10.1214/ss/1177011136.

Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 6 2000. doi: https://doi.org/10.1161/01.CIR.101.23.e215.

Bruno Grassi, Harry B. Rossiter, and Jerzy A. Zoladz. Skeletal muscle fatigue and decreased efficiency: Two sides of the same coin? *Exercise and Sport Sciences Reviews*, 43(2):75–83, 4 2015. doi: https://doi.org/10.1249/JES.0000000000000043.

Siddhant Gupta, Ankur T. Patil, Mirali Purohit, Mihir Parmar, Maitreya Patel, Hemant A. Patil, and Rodrigo Capobianco Guido. Residual neural network precisely quantifies dysarthria severity-level based on short-duration speech segments. *Neural Networks*, 139:105–117, 7 2021. doi: https://doi.org/10.1016/j.neunet.2021.02.008.

Jean Hausser and Korbinian Strimmer. Entropy inference and the james-stein estimator, with application to nonlinear gene association networks. *The Journal of Machine Learning Research*, 10:1469–1484, 12 2009.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

Matthew D. Homan and Andrew Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *The Journal of Machine Learning Research*, 15(1):1593–1623, 1 2014.

Martin Jankowiak and Fritz Obermeyer. Pathwise derivatives beyond the reparameterization trick. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2235–2244. PMLR, 7 2018.

Denis Jordan, Gudrun Stockmanns, Eberhard F. Kochs, Stefanie Pilge, and Gerhard Schneider. Electroencephalographic order pattern analysis for the separation of consciousness and unconsciousness: An analysis of approximate entropy, permutation entropy, recurrence rate, and phase coupling of order recurrence plots. *Anesthesiology*, 109(6): 1014–1022, 12 2008. doi: https://doi.org/10.1097/ALN.0b013e31818d6c55.

Sadeem Nabeel Saleem Kbah, Noor Kamal Al-Qazzaz, Sumai Hamad Jaafer, and Mohannad K. Sabir. Epileptic eeg activity detection for children using entropy-based biomarkers. *Neuroscience Informatics*, 2(4):100101, 12 2022. doi: https://doi.org/10.1016/j.neuri.2022.100101.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations*, 2015. doi: https://doi.org/10.48550/arXiv.1412.6980.

Karin C Knudson and Jonathan W Pillow. Spike train entropy-rate estimation using hierarchical dirichlet process priors. In *Advances in Neural Information Processing Systems*, volume 26, pages 2076–2084, 12 2013.

Douglas J. Little, Joshua P. Toomey, and Deb M. Kane. Efficient bayesian estimation of permutation entropy with dirichlet priors. *Communications in Nonlinear Science and Numerical Simulation*, 108:106216, 5 2022. doi: https://doi.org/10.1016/j.cnsns.2021.106216.

Daniel McNeish. On using bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(5): 750–773, 6 2016. doi: https://doi.org/10.1080/10705511.2016.1186549.

Ilya Nemenman, F. Shafee, and William Bialek. Entropy and inference, revisited. In *Advances in Neural Information Processing Systems*, volume 14, pages 471–478. MIT Press, 2001.

E. Olofsen, J.W. Sleigh, and A. Dahan. Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. *British Journal of Anaesthesia*, 101(6):810–821, 12 2008. doi: https://doi.org/10.1093/bja/aen290.

Liam Paninski. Estimation of entropy and mutual information. *Neural Computation*, 15(6): 1191–1253, 6 2003. doi: https://doi.org/10.1162/089976603321780272.

Omiros Papaspiliopoulos, Gareth O. Roberts, and Martin Sköld. A General Framework for the Parametrization of Hierarchical Models. *Statistical Science*, 22(1):59 – 73, 2 2007. doi: https://doi.org/10.1214/088342307000000014.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2 2019.

Fernando E. Pose, Lucas Bautista, Franco Gianmuso, and Francisco O. Redelico. On the permutation entropy bayesian estimation. *Communications in Nonlinear Science and Numerical Simulation*, 99: 105779, 8 2021. doi: https://doi.org/10.1016/j.cnsns.2021.105779.

Jee S. Ra, Tianning Li, and Yan Li. A novel permutation entropy-based eeg channel selection for improving epileptic seizure prediction. *Sensors*, 21(23), 11 2021. doi: https://doi.org/10.3390/s21237972.

Jagdeep Rahul and Lakhan Dev Sharma. Artificial intelligence-based approach for atrial fibrillation detection using normalised and short-duration time-frequency ecg. *Biomedical Signal Processing and Control*, 71:103270, 1 2022. doi: https://doi.org/10.1016/j.bspc.2021.103270.

Jagdeep Rahul, Lakhan Dev Sharma, and Vijay Kumar Bohat. Short duration vectorcardiogram based inferior myocardial infarction detection: class and subject-oriented approach. *Biomedical Engineering / Biomedizinische Technik*, 66(5):489–501, 5 2021. doi: https://doi.org/10.1515/bmt-2020-0329.

Leonardo Ricci and Alessio Perinelli. Estimating permutation entropy variability via surrogate time series. *Entropy*, 24(7), 6 2022. doi: https://doi.org/10.3390/e24070853.

Maik Riedl, Andreas Müller, and Niels Wessel. Practical considerations of permutation entropy. *Eur. Phys. J. Spec. Top.*, 222:249–262, 6 2013. doi: https://doi.org/10.1140/epjst/e2013-01862-7.

Zbyněk Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62 (318):626–633, 6 1967. doi: https://doi.org/10.2307/2283989.

Francisco Traversaro and Francisco O Redelico. Confidence intervals and hypothesis testing for the permutation entropy with an application to epilepsy. *Communications in Nonlinear Science and Numerical Simulation*, 57:388–401, 4 2018. doi: https://doi.org/10.1016/j.cnsns.2017.10.013.

Iman Veisi, Naser Pariz, and Ali Karimpour. Fast and robust detection of epilepsy in noisy eeg signals using permutation entropy. In *2007 IEEE 7th International Symposium on BioInformatics and BioEngineering*, pages 200–203, 2007. doi: https://doi.org/10.1109/BIBE.2007.4375565.

Karlman Wasserman, James E. Hansen, Daryl Y. Sue, Brian J. Whipp, and Victor F. Froelicher. Principles of exercise testing and interpretation. *Journal of Cardiopulmonary Rehabilitation*, 7(4): 189, 4 1987.

Jian Yin, PengXiang Xiao, Junyan Li, Yungang Liu, Chenggang Yan, and Yatao Zhang. Parameters analysis of sample entropy, permutation entropy and permutation ratio entropy for rr interval time series. *Information Processing & Management*, 57 (5):102283, 9 2020. doi: https://doi.org/10.1016/j.ipm.2020.102283.

Robert Zanner, Sebastian Berger, Natalie Schröder, Matthias Kreuzer, and Gerhard Schneider. Separation of responsive and unresponsive patients under clinical conditions: comparison of symbolic transfer entropy and permutation entropy. *Journal of Clinical Monitoring and Computing*, 38(1): 187–196, 7 2023. doi: https://doi.org/10.1007/s10877-023-01046-w.

Lina Zhao, Shoushui Wei, Chengqiu Zhang, Yatao Zhang, Xinge Jiang, Feng Liu, and Chengyu Liu. Determination of sample entropy and fuzzy measure entropy parameters for distinguishing congestive heart failure from normal sinus rhythm subjects. *Entropy*, 17(9):6270–6288, 2015. doi: https://doi.org/10.3390/e17096270.

Appendix A. Permutation Symbol Visualization

Figure 4 illustrates the full set of permutation symbols for an embedding dimension of $m = 3$.

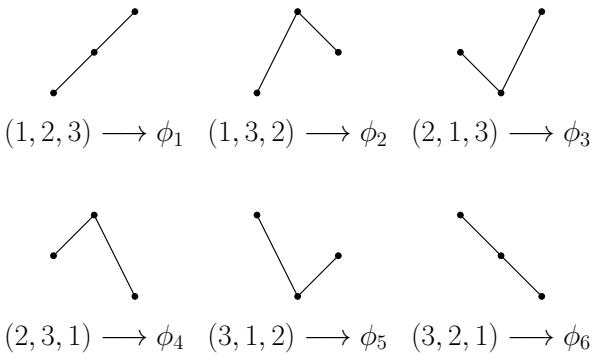

Figure 4: Complete set of permutation symbol mappings for $m = 3$ embedding dimension size.

Each symbol is derived using a mapping function, g, which takes a rank-ordered vector (e.g., $(1, 2, 3)$) as input and assigns it to a unique permutation symbol. Given a rank-ordered vector of size m, there are precisely $m!$ distinct permutation symbols, with each rank-ordered vector mapping one-to-one to a permutation symbol. The rank-ordering reflects the relative positioning within the original sub-vector; for example, the vector $(1, 3, 2)$ indicates that the first element was the smallest, the second was the largest, and the third was intermediate in value.

We assume that no ties exist between values in a sub-vector, a condition that is practically guaranteed when the original signal, $\boldsymbol{x}$, consists of continuous,

real-valued numbers. To ensure this assumption is met, small perturbations may be introduced into the signal.

Appendix B. Gamma Hyperprior Optimization Numerical Tricks

To construct a broader class of implied priors for our Bayesian PermEn estimator, we model the concentration hyperprior as $\alpha \sim \Gamma(a, b)$, and manipulate (a, b) to minimize the W_1 distance between $\mathcal{H}_m$ and a reference measure η. The principle challenge under this framework is to obtain estimates for the gradients: $\frac{\partial W_1}{\partial a}$ and $\frac{\partial W_1}{\partial b}$, for the optimization routine.

Under the proposed non-centered model detailed in Equation (5), the W_1 distance error gradient with respect to a is calculated as $\frac{\partial W_1}{\partial a} = \frac{\partial W_1}{\partial \mathcal{H}_m} \frac{\partial \mathcal{H}_m}{\partial \boldsymbol{\pi}} \frac{\partial \boldsymbol{\pi}}{\partial \widetilde{\boldsymbol{\pi}}} \frac{\widetilde{\boldsymbol{\pi}}}{\partial \alpha} \frac{\partial \alpha}{\partial a}$, and similarly for b. The difficulty is that $\frac{\partial \alpha}{\partial a}$ does not admit a simple analytic expression and thus we must resort to numerical approximations of this gradient.

Fortunately, it is equivalent to express the hyperprior as $\frac{\alpha}{b} \sim \Gamma(a, 1)$, and it suffices to only consider the derivative, $\frac{d\alpha}{da}$. Although $\frac{\alpha}{b}$ is a stochastic node, Jankowiak and Obermeyer (2018) have devised a deterministic Taylor series approximation for $\frac{d\alpha}{da}$, which exhibits a relative accuracy of 0.0005 across a wide range of inputs. This approximate pathwise gradient enables us to compute $\frac{\partial W_1}{\partial a}$ and, consequently, $\frac{\partial W_1}{\partial b}$, as all remaining operations are deterministic and differentiable

We implemented our approach using the PyTorch framework (Paszke et al., 2019) and employed the stochastic gradient optimizer, Adam (Kingma and Ba, 2015), with a learning rate of $\lambda = 0.01$. We generated $10,000$ samples for both η and $\mathcal{H}_m$, and ran the optimization algorithm for 1000 epochs to ensure convergence.

Appendix C. NSB vs Optimized Beta(1, 1) PermEn Implied Priors

In Figure 2, we observed an unexpected result: for both the AR(1) and PAF signals, the NSB estimator had higher error compared to our Logit(1, 1) estimator. This finding contradicted our initial expectations since the hierarchical prior defined by Equation (3) approximately implies a uniform prior over $[0, \log m!]$ (or, after a rescaling, over $[0, 1]$) – the same implied prior for the Logit(1, 1) model. We hypothesize that this difference arises from the superior numerical properties of the reference priors generated by our algorithm relative to the implied process proposed by Nemenman et al. (2001).

To test this hypothesis, we conducted an analysis using MCMC sampling. We generated instances of concentration priors following Equation (3) and employed the Dirichlet-Multinomial framework to sample implied PermEn prior distributions under this model. Additionally, we used our optimization algorithm to minimize the implied PermEn prior while using a reference prior of $\eta \sim \text{Beta}(1, 1)$. This process was repeated 25 times with different random seeds for both approaches. We then calculated the Wasserstein-1 distance between the implied PermEn prior and a uniform prior over the interval $[0, 1]$, and compared the relative errors between the estimators using a one-sided, non-equal variance t-test with the alternative hypothesis being that our method had lower Wasserstein error. The results of this analysis are displayed in Table 2.

Table 2: Our proposed optimization scheme under an $\eta \sim \text{Beta}(1, 1)$ reference prior achieves significantly lower relative error than NSB hierarchical prior at all embedding dimension sizes.

m	NSB	Logit(1, 1)	P-value
3	0.037 ± 0.003	0.012 ± 0.0003	6.28×10^{-25}
4	0.017 ± 0.002	0.0055 ± 0.0002	1.28×10^{-19}
5	0.012 ± 0.002	0.0047 ± 0.0002	8.57×10^{-15}
6	0.010 ± 0.002	0.0053 ± 0.0003	2.42×10^{-10}

At all the standard embedding dimension sizes for PermEn ($m \in \{3, 4, 5, 6\}$), our logit-based approach consistently exhibits lower relative Wasserstein-1 error for the reference prior of $\eta \sim \text{Beta}(1, 1)$ compared to the NSB hierarchical concentration prior. This observed effect is statistically significant for all values. The key factor contributing to this outcome is the tendency of the NSB approach to relatively overweight both the upper and lower boundaries of the PermEn distribution in comparison to our approach. For evidence of this phenomenon, observe the probability density functions displayed in Figure 5.

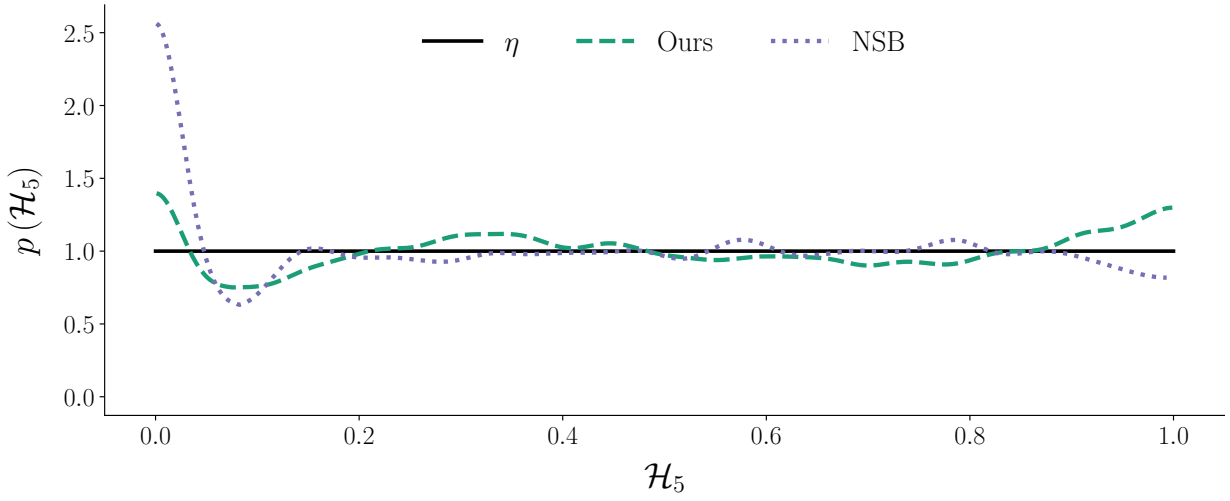

Figure 5: The NSB estimator puts too much relative weight on the left tail of the Beta(1, 1) distribution causing the resulting posterior PermEn estimate to favor lower values.

Appendix D. PermEn Posterior Inference Computational Time Analysis

To evaluate the computational efficiency of our proposed PermEn estimator against existing methods, we analyzed the time required to approximate the PermEn posterior distribution. This comparison included our estimator, the Perks' prior, an NSB estimator calculating only the first and second moments of the posterior distribution, and a full NSB posterior approximation using MCMC sampling methods. For the NSB estimator's full posterior approximation, we used the the Metropolis-Hastings algorithm due to the absence of a straightforward sampling method for the kernel specified in Equation (3) and the trigamma function's unstable gradients. While valid, the Metropolis-Hastings algorithm is generally less efficient than Hamiltonian Monte Carlo methods, such as the No U-Turn Sampler (Homan and Gelman, 2014), due to its tendency to perform a random walk around the typical set of the posterior without leveraging gradient information to efficiently explore the parameter space (Betancourt, 2018).

Our computational experiments aimed to assess how embedding dimension (m) and signal length (N) affect the time required to infer the posterior PermEn. To simulate count vectors, we generated probability vectors ($\boldsymbol{\pi}$) from a Dirichlet distribution (Dir(1)), which were then used to produce multinomial-distributed count vectors ($\boldsymbol{c} \sim \text{Mult}(N, \boldsymbol{\pi})$). We fixed the signal length at $N = 25$ and varied the embedding dimension ($m \in \{3, 4, 5, 6\}$) for the embedding dimension impact experiments. Conversely, we set $m = 5$ and varied signal length ($N \in \{10, 25, 50\}$) for the signal length impact experiments. Each configuration was repeated 20 times to calculate the median execution time, displayed in Figure 6.

We observed a more pronounced computational impact from increasing the embedding dimension compared to increasing the signal length. The NSB estimator, which computes only the posterior's first and second moments, demonstrated superior efficiency for $m > 3$, likely due to optimized numerical algorithms for one-dimensional integration. However, for full posterior approximation with the NSB prior at $m = 3$ and $m = 4$, it took almost an order of magnitude more computation time to converge than our proposed estimator. For $m \geq 5$, the MCMC sampler failed to converge after over two million draws, in contrast to our estimator which required approximately 10,000 draws. The Perks' prior, employing a Bayesian conjugate model, was inherently more efficient for approximating the complete posterior, but exhibited poorer scaling with increasing m, reflect-

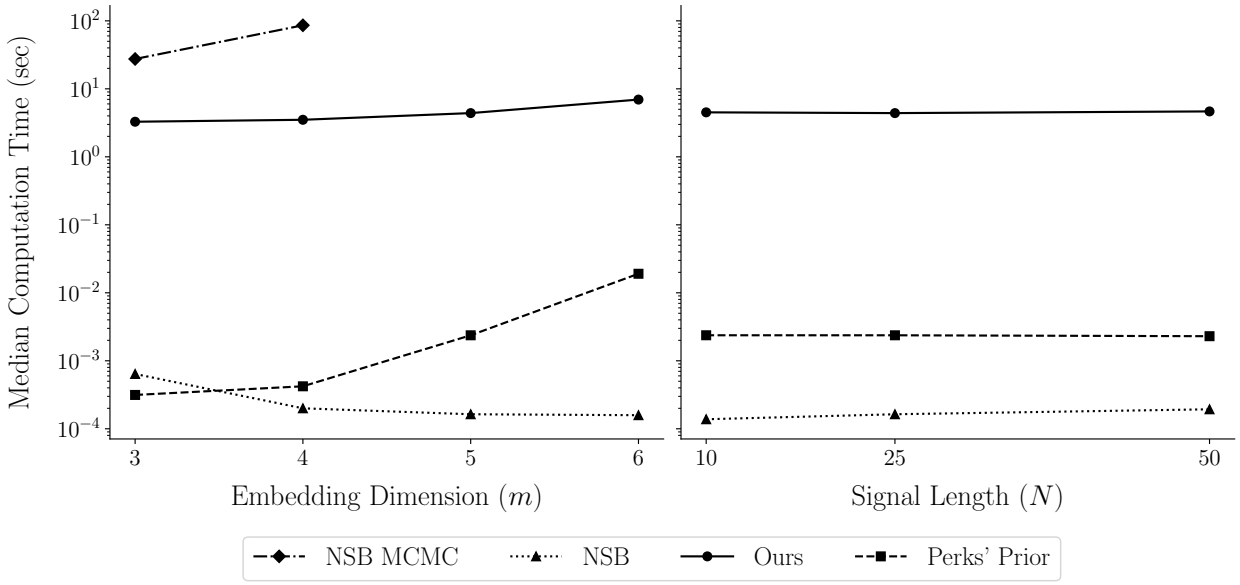

Figure 6: All methods except full MCMC NSB can infer the posterior PermEn in a reasonable computational timeframe. NSB estimating the first and second moments is the most efficient, particularly for higher embedding dimension sizes.

ing the computational challenge of sampling from a high-dimensional probability simplex.

The influence of signal length (N), at a fixed embedding dimension ($m = 5$), was relatively minor. While the NSB estimator showed a slight increase in computational time, our model and the Perks' prior experienced negligible changes. This suggests that for our model, the dominant computational factors were the compilation of sampling and gradient estimation functions, rather than the length of the signal itself. A substantially longer signal might be necessary to significantly influence the computational demand, as the time required for MCMC sampling is generally a linear function of the number of observations and MCMC steps (Belloni and Chernozhukov, 2009).