

Minimizing the Expected Posterior Entropy Yields Optimal Summary Statistics

Till Hoffmann

Harvard T.H. Chan School of Public Health, US.

29th September 2022, 1.30pm UK time

Abstract

Extracting low-dimensional summary statistics from large datasets is essential for efficient (likelihood-free) inference. We propose obtaining summary statistics by minimizing the expected posterior entropy (EPE) under the prior predictive distribution of the model. We show that minimizing the EPE is equivalent to learning a conditional density estimator for the posterior as well as other information-theoretic approaches. Further summary extraction methods (including minimizing the L^2 Bayes risk, maximizing the Fisher information, and model selection approaches) are special or limiting cases of EPE minimization. We demonstrate that the approach yields high fidelity summary statistics by applying it to both a synthetic benchmark as well as a population genetics problem. We not only offer concrete recommendations for practitioners but also provide a unifying perspective for obtaining informative summary statistics.

Reference: T. Hoffmann and J.P. Onnela. Minimizing the Expected Posterior Entropy Yields Optimal Summary Statistics. Preprint at ArXiv:2206.02340, 2022.