

An offline sampling design problem for distributed detection is considered in this paper. To reduce the sensing, storage, transmission, and processing costs, the natural choice for the sampler is the sparsest one that results in a desired global error probability. Since the numerical optimization of the error probabilities is difficult, we adopt simpler costs related to distance measures between the conditional distributions of the sensor observations. We design sparse samplers for the Bayesian as well as the Neyman-Pearson setting. The developed theory can be applied to sensor placement/selection, sample selection, and fully decentralized data compression. For conditionally independent observations, we give an explicit solution, which is optimal in terms of the error exponents. More specifically, the best subset of sensors is the one with the smallest local average root-likelihood ratio and largest local average log-likelihood ratio in the Bayesian and Neyman-Pearson setting, respectively. We supplement the proposed framework with a thorough analysis for Gaussian observations, including the case when the sensors are conditionally dependent, and also provide examples for other observation distributions. One of the results shows that, for nonidentical Gaussian sensor observations with uncommon means and common covariances under both hypotheses, the number of sensors required to achieve a desired detection performance reduces significantly as the sensors become more coherent.