Abstract: Statistical distances, such as the KL divergence and Wasserstein distances, are powerful tools for measuring differences between probability distributions. Unfortunately, the sample complexity of estimating these distances often scales prohibitively poorly with increasing dimension. In this work, we study the convergence of the empirical measure smoothed by a Gaussian kernel to the true measure convolved with a Gaussian, as measured by various statistical distances. We show that these distances converge at the parametric rate under certain conditions. We consider applications of these smoothed statistical distances to high dimensional settings in machine learning, including estimation of entropy and mutual information in neural networks (e.g. information bottleneck), statistical-distance based estimation of distributions from samples, and generative modeling.