

“ONLINE” QUANTILE AND DENSITY ESTIMATORS

Raghu Pasupathy, Virginia Tech, VA, USA, pasupath@vt.edu
Soumyadip Ghosh, IBM Research, NY, USA, ghoshs@us.ibm.com

The standard estimator $\hat{q}_\alpha(n)$ for the α -quantile q_α of a random variable X , given n observations from the distribution of X , is obtained by inverting the empirical cumulative distribution function (cdf) constructed from the obtained observations. It is well-known that $\hat{q}_\alpha(n)$ requires $O(n)$ storage, and that the mean squared error of $\hat{q}_\alpha(n)$ (with respect to q_α) decays as $O(n^{-1})$. In this talk, we present an alternative to $\hat{q}_\alpha(n)$ that seems to require dramatically less storage with negligible loss in convergence rate. The proposed estimator, $\tilde{q}_\alpha(n)$, relies on an alternative cdf that is constructed by accumulating the observed random variates into variable-sized bins that progressively become finer around the quantile. The size of the bins are strategically adjusted to ensure that the increased bias due to binning does not adversely affect the resulting convergence rate. We will present an “online” version of the estimator $\tilde{q}_\alpha(n)$, along with results on its consistency, convergence rates, and storage requirements. If time permits, we will discuss analogous ideas for density estimation.