Information Criterion for Minimum Cross-Entropy Model Selection

Youngjun Choe, PhD Department of Industrial & Systems Engineering University of Washington

Abstract: The cross-entropy, which is proportional to the Kullback-Leibler divergence, is widely used to gauge the deviation of a distribution of interest from a reference distribution in statistical inference. For example, the Akaike information criterion (AIC) is an asymptotically unbiased estimator of the cross-entropy from a parametric distribution to the true distribution of data. Minimizing the AIC allows us to find a parametric distribution close to the true distribution. In this paper, we generalize the AIC by letting the reference distribution be a target distribution to approximate when its density can be evaluated up to a multiplicative constant only at observed data points. We prove, under some conditions, that the generalized criterion, which we call the cross-entropy information criterion (CIC), is an asymptotically unbiased estimator of the target distribution. We demonstrate the usefulness of CIC for approximating the optimal importance sampling distribution by a mixture of parametric distributions.

The preprint of this paper is available at <u>https://arxiv.org/abs/1704.04315</u>. This talk will be generally accessible to ISE graduate students. Connections of this work to optimization and Bayesian inference (through importance sampling) will be briefly discussed as well.

Bio: *Youngjun Choe* is an Assistant Professor of Industrial & Systems Engineering at the University of Washington, Seattle. His research centers around developing statistical methods to infer on extreme events (e.g., wind turbine failures and power outages) using empirical and simulated data. He received his Ph.D. in Industrial & Operations Engineering and M.A. in Statistics from the University of Michigan, Ann Arbor. He holds bachelor's degrees in Physics and Management Science from KAIST in Korea.

Tuesday, December 5, 2017 1:30 – 2:20 p.m. MEB 235