

Abstract

The information content, or equivalently, the entropy score is a widely used criterion to score a putative sequence motif. Efficient and accurate evaluation of the statistical significance of such a score is an essential requirement for motif-finding tools. A common approach to evaluate this significance is to compute the p -value of the observed score. Current computation schemes used in popular motif-finding programs can unwittingly provide poor approximations of this p -value. We present an approach to a fast and reliable estimation. Our technique, which is backed up by rigorous error analysis can be applied in a more general setting.