Discrete MDL Predicts in Total Variation

Marcus Hutter, ANU&NICTA, Canberra, Australia, www.hutter1.net

Main result informal: For *any* countable class of models $\mathcal{M} = \{Q_1, Q_2, ...\}$ containing the unknown true sampling distribution P, MDL predictions converge to the true distribution in total variation distance. Formally ...

- Given $x = x_1...x_\ell$, the Q-prob. of $z = x_{\ell+1}x_{\ell+2}...$ is $Q(z|x) = \frac{Q(xz)}{Q(x)}$
- Use $Q =$ Bayes or $Q =$ MDL instead of P for prediction
- Total variation distance: $d_\infty(P,Q) := \sup_{A\subseteq\mathcal{X}^\infty}\big|Q[A|x] - P[A|x]\big|$
- Bayes(x) $:= \sum_{Q\in\mathcal{M}} Q(x) w_Q$, $\qquad$ [$w_Q > 0\, \forall Q \in \mathcal{M}$ and $\sum_{Q\in\mathcal{M}} w_Q = 1$]
- MDL selects Q which leads to minimal code length for x:
  MDL^x $:= \arg\min_{Q\in\mathcal{M}}\{-\log Q(x) + K(Q)\}$, $\qquad$ [$\sum_{Q\in\mathcal{M}} 2^{-K(Q)} \leq 1$]

Theorem 1 (Discrete Bayes&MDL Predict in Total Variation)

$$d_\infty(P, \text{Bayes}|x) \to 0 \qquad \left\{\begin{array}{l}\text{almost surely}\\ \text{for } \ell(x)\to\infty\end{array}\right\} \qquad \text{[Blackwell\&Dubins 1962]}$$

$$d_\infty(P, \text{MDL}^x|x) \to 0 \qquad \text{[Hutter NIPS 2009]}$$

No *independence*, *ergodicity*, *stationarity*, *identifiability*, or other assumption