

**Definition 2.23** Let  $p$  and  $q$  be two probability distributions on a common alphabet  $\mathcal{X}$ . The variational distance between  $p$  and  $q$  is defined as

$$V(p, q) = \sum_{x \in \mathcal{X}} |p(x) - q(x)|.$$

The entropy function is continuous at  $p$  if

$$\lim_{p' \rightarrow p} H(p') = H \left( \lim_{p' \rightarrow p} p' \right) = H(p),$$

or equivalently, for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that

$$|H(p) - H(q)| < \epsilon$$

for all  $q \in \mathcal{P}_{\mathcal{X}}$  satisfying

$$V(p, q) < \delta.$$