

**Definition 2.28** The informational divergence between two probability distributions  $p$  and  $q$  on a common alphabet  $\mathcal{X}$  is defined as

$$D(\textcolor{red}{p}||\textcolor{blue}{q}) = \sum_x \textcolor{red}{p}(\textcolor{red}{x}) \log \frac{\textcolor{red}{p}(\textcolor{red}{x})}{\textcolor{blue}{q}(\textcolor{blue}{x})} = E_{\textcolor{red}{p}} \log \frac{\textcolor{red}{p}(\textcolor{red}{X})}{\textcolor{blue}{q}(\textcolor{blue}{X})},$$

where  $E_{\textcolor{red}{p}}$  denotes expectation with respect to  $\textcolor{red}{p}$ .