Examples

Classical: $\Phi(P,\kappa) = \langle \kappa, P\rangle = \sum p_i\kappa_i$, P a probability distribution, κ a code (i.e. Krafts equality applies: $\sum \exp(-\kappa_i) = 1$). Gives classical entropy and divergence. Map $P \curvearrowright \widehat{P} = \kappa$ given by $\kappa_i = \ln\frac{1}{p_i}$. May avoid $Y \neq X$. Then $\Phi(P,Q) = \sum p_i \ln\frac{1}{q_i}$, $D(P\|Q) = \sum p_i \ln\frac{p_i}{q_i}$.

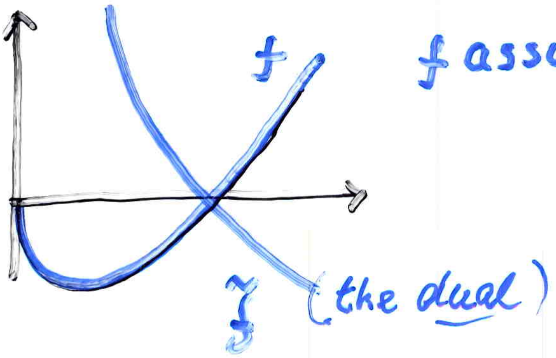
via *Csiszár divergences*:

$$\Phi = \sum_{i\in\mathbb{A}} \Big(q_i f(\frac{p_i}{q_i}) - f(p_i)\Big) = \sum_{i\in\mathbb{A}} p_i\Big(\tilde f\big(\frac{q_i}{p_i}\big) - \tilde f\big(\frac{1}{p_i}\big)\Big)$$

$$H = \sum_{i\in\mathbb{A}} \Big(-f(p_i)\Big) = \sum_{i\in\mathbb{A}} \Big(-p_i\tilde f\big(\frac{1}{p_i}\big)\Big)$$

$$\mathrm{D} = \sum_{i\in\mathbb{A}} q_i f(\frac{p_i}{q_i}) = \sum_{i\in\mathbb{A}} p_i \tilde f\big(\frac{q_i}{p_i}\big)$$

(Classical: $f(x) = x\ln(x)$, $\tilde f(x) \stackrel{\triangle}{=} x f(\frac{1}{x}) = \ln\frac{1}{x}$)



f assumed convex, $f(1)=0$.