

Pinsker's inequality with a weaker constant

A proof using the log inequality and Cauchy–Schwarz

Lecture 4 supplement · 5 October 2026

Let P, Q be probability distributions on a finite alphabet $\mathcal{Z}$. Use base-two logarithms for KL divergence and the half- L^1 convention for TV:

$$D_2(P\|Q) = \sum_{z:P(z)>0} P(z) \log_2 \frac{P(z)}{Q(z)}, \quad \text{TV}(P, Q) = \frac{1}{2} \sum_z |P(z) - Q(z)|.$$

We will prove

$$\boxed{\text{TV}(P, Q)^2 \leq (\ln 2) D_2(P\|Q)}.$$

The idea is to compare both quantities with the squared difference of their square roots. The sharp Pinsker bound has $\ln 2/2$ in place of $\ln 2$.

Proof. If $D_2(P\|Q) = \infty$, the inequality is immediate. Otherwise $Q(z) > 0$ whenever $P(z) > 0$. Define

$$S = \sum_z (\sqrt{P(z)} - \sqrt{Q(z)})^2 = 2 - 2 \sum_z \sqrt{P(z)Q(z)}.$$

1. KL controls S . Apply the elementary inequality $-\ln t \geq 1 - t$ for $t > 0$:

$$\begin{aligned} (\ln 2) D_2(P\|Q) &= -2 \sum_{z:P(z)>0} P(z) \ln \sqrt{\frac{Q(z)}{P(z)}} \\ &\geq 2 \sum_{z:P(z)>0} P(z) \left(1 - \sqrt{\frac{Q(z)}{P(z)}}\right) \\ &= 2 - 2 \sum_z \sqrt{P(z)Q(z)} = S. \end{aligned}$$

Terms with $P(z) = 0$ contribute zero to the last sum.

2. S controls TV. Factor $P(z) - Q(z)$ as a difference of squares and use Cauchy–Schwarz:

$$\begin{aligned} \text{TV}(P, Q)^2 &= \frac{1}{4} \left[\sum_z |\sqrt{P(z)} - \sqrt{Q(z)}| (\sqrt{P(z)} + \sqrt{Q(z)}) \right]^2 \\ &\leq \frac{1}{4} S \sum_z (\sqrt{P(z)} + \sqrt{Q(z)})^2 \leq S. \end{aligned}$$

For the last step, Cauchy–Schwarz gives $\sum_z \sqrt{P(z)Q(z)} \leq 1$, because both distributions sum to one. Hence the remaining sum is $2 + 2 \sum_z \sqrt{P(z)Q(z)} \leq 4$. Combining the two steps proves the claim. $\square$

Binary case. For $P = \text{Bernoulli}(p)$ and $Q = \text{Bernoulli}(q)$, TV is $|p - q|$, so

$$d_2(p\|q) \geq \frac{(p - q)^2}{\ln 2}.$$

Average encoding. For any finite jointly distributed (X, M) , apply the proved bound to P_{XM} and $P_X P_M$. The joint TV expands as the mean conditional TV, giving

$$\mathbb{E}_X [\text{TV}(P_{M|X}, P_M)] = \text{TV}(P_{XM}, P_X P_M) \leq \sqrt{(\ln 2) I(X; M)}.$$

This bound is weaker than sharp Pinsker by a factor of $\sqrt{2}$ in TV.

Source and further reading. For the related Hellinger-affinity argument, see Clément L. Canonne, *A short note on an inequality between KL and TV*, arXiv:2202.07198v2 (2023), Section 2, equations (8)–(9). Canonne credits Tsybakov's Lemma 2.6 for that argument. The proof above is an elementary variant using $-\ln t \geq 1 - t$.