This proof of Pinsker's lemma is due to Pollard: `http://www.stat.yale.edu/~pollard/Books/Asymptopia/Metrics.pdf`.

Let P, Q be two distributions over some set Ω. We'll treat Ω as a discrete set but all of this carries over to the continuous case as well. Denote by p, q the probability mass functions of P, Q, respectively.

We have already encountered the **total variation distance**

$$\frac{1}{2}\left\|P-Q\right\|_1 = \frac{1}{2}\sum_{x\in\Omega}|p(x)-q(x)|.$$

Define the **Kullback-Leibler divergence** $\mathrm{KL}(P||Q)$:

$$\mathrm{KL}(P||Q) = \sum_{x\in\Omega} p(x)\log\frac{p(x)}{q(x)} = \mathbb{E}_{X\sim P}\left[\log\frac{p(X)}{q(X)}\right].$$

Pinsker's inequality relates these two quantities:

Theorem 0.1

$$\mathrm{KL}(P||Q) \ge \frac{1}{2}\left\|P-Q\right\|_1^2.$$

Proof: The following simple fact is left as an exercise:

$$(1+t)\log(1+t) - t \ge \frac{1}{2}\cdot\frac{t^2}{1+t/3}, \qquad t \ge -1. \tag{1}$$

There is no loss of generality in assuming that

$$\sup_{x\in\Omega}\frac{p(x)}{q(x)} < \infty, \tag{2}$$

since otherwise $\mathrm{KL}(P||Q) = \infty$ and the claim is vacuously true.

Assuming (2), we may take $p(x) = (1+r(x))q(x)$ where $r(x) = p(x)/q(x) - 1 \ge -1$.[1]

Exercise: verify that

$$\mathbb{E}_{X\sim Q}[r(X)] = \sum_{x\in\Omega} q(x)r(x) = 0, \tag{3}$$

$$\mathbb{E}_{X\sim Q}[|r(X)|] = \sum_{x\in\Omega} q(x)|r(x)| = \left\|P-Q\right\|_1, \tag{4}$$

$$\mathrm{KL}(P||Q) = \mathbb{E}_{X\sim Q}[(1+r(X))\log(1+r(X)) - r(X)]. \tag{5}$$

[1]We define $0/0 = 1$.

Combining (1) and (5), we get

$$\mathrm{KL}(P||Q) \quad \geq \quad \frac{1}{2}\mathbb{E}_{X\sim Q}\left[\frac{r(X)^2}{1+r(X)/3}\right].$$

Now (3) implies $\mathbb{E}_Q[1+r(X)/3] = 1$, and hence

$$\mathrm{KL}(P||Q) \quad \geq \quad \frac{1}{2}\mathbb{E}_Q\left[\frac{r(X)^2}{1+r(X)/3}\right]\mathbb{E}_Q[1+r(X)/3].$$

Recall the Cauchy-Schwarz inequality:

$$(\mathbb{E}_Q[f(X)g(X)])^2 \leq \mathbb{E}_Q[f(X)^2]\mathbb{E}_Q[g(X)^2].$$

Taking $f(x) = \sqrt{r(x)^2/(1+r(x)/3)}$ and $g(x) = \sqrt{1+r(x)/3}$, we have

$$\begin{aligned}
\mathrm{KL}(P||Q) \quad &\geq \quad \frac{1}{2}\mathbb{E}_Q[f(X)^2]\mathbb{E}_Q[g(X)^2] \\
&\geq \quad \frac{1}{2}\left(\mathbb{E}_Q[f(X)g(X)]\right)^2 \\
&= \quad \frac{1}{2}\left(\mathbb{E}_Q\left[\frac{|r(X)|}{\sqrt{1+r(X)^2}}\cdot\sqrt{1+r(X)^2}\right]\right)^2 \\
&= \quad \frac{1}{2}\left(\mathbb{E}_Q\left[|r(X)|\right]\right)^2 \\
&= \quad \frac{1}{2}\left\|P-Q\right\|_1^2,
\end{aligned}$$

where the last equality follows from (4). ■