556: MATHEMATICAL STATISTICS I

THE KULLBACK-LEIBLER DIVERGENCE

The **Kullback-Leibler** (KL) **divergence** between two probability distributions with densities f_0 and f_1 with supports $\mathbb{X}_0$ and $\mathbb{X}_1$ respectively is defined as

$$\mathbb{K}\left(f_0, f_1\right) = \int_{\mathbb{X}_0} \log \frac{f_0(x)}{f_1(x)} f_0(x)\,dx = \mathbb{E}_{f_0}\left[\log \frac{f_0(X)}{f_1(X)}\right]$$

Using **Jensen's Inequality** on the convex function $g(x) = -\log x$

$$\begin{aligned}
-\mathbb{K}\left(f_0, f_1\right) = \mathbb{E}_{f_0}\left[-\log \frac{f_0(X)}{f_1(X)}\right] = \mathbb{E}_{f_0}\left[\log \frac{f_1(X)}{f_0(X)}\right] &\leq \log \mathbb{E}_{f_0}\left[\frac{f_1(X)}{f_0(X)}\right] \\
&= \log\left\{\int_{\mathbb{X}_0} \frac{f_1(x)}{f_0(x)} f_0(x)\,dx\right\} \\
&= \log\left\{\int_{\mathbb{X}_0} f_1(x)\,dx\right\} \leq \log 1 = 0
\end{aligned}$$

Hence

$$\mathbb{K}\left(f_0, f_1\right) \geq 0.$$

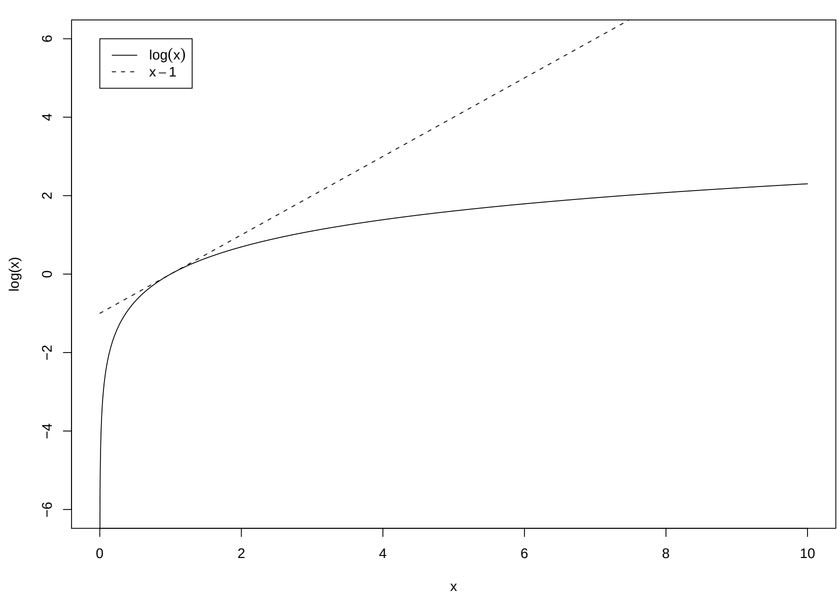
Now clearly, if f_0 and f_1 are identical, so that $f_1(x) = f_0(x)$ for all $x \in \mathbb{X}_0 \equiv \mathbb{X}_1$, then

$$\mathbb{K}\left(f_0, f_1\right) = 0.$$

For the converse, note that for all real $x \geq 0$

$$\log x \leq x - 1 \tag{1}$$

with equality only when $x = 1$, as the plot demonstrates.



Therefore

$$\log \frac{f_0(x)}{f_1(x)} \leq \frac{f_0(x)}{f_1(x)} - 1$$

with equality only if for all x

$$\frac{f_0(x)}{f_1(x)} = 1 \qquad \text{or} \qquad f_1(x) = f_0(x)$$

In the KL calculation, only if $f_1(x) = f_0(x)$,

$$\mathbb{E}_{f_0}\left[\log \frac{f_1(x)}{f_0(x)}\right] = \mathbb{E}_{f_0}\left[\left(\frac{f_1(x)}{f_0(x)} - 1\right)\right],$$

so therefore

$$\mathbb{E}_{f_0}\left[\left(\frac{f_1(X)}{f_0(X)} - 1\right)\right] = \int \left(\frac{f_1(x)}{f_0(x)} - 1\right) f_0(x)\, dx = \int \left(f_1(x) - f_0(x)\right)\, dx = 0$$

only if $f_1(x) = f_0(x)$.

Therefore

$$\mathbb{K}(f_0, f_1) = 0 \qquad \Longleftrightarrow \qquad f_1(x) = f_0(x) \quad \text{for all } x \in \mathbb{X}_0 \equiv \mathbb{X}_1.$$

Exercise: Prove equation (1) without the use of graphical aids; show that

$$x - 1 - \log x \geq 0$$

for all x.

Note that

$$\mathbb{K}(f_0, f_1) \neq \mathbb{K}(f_1, f_0)$$

so the **divergence** is not a **distance measure** as it is not symmetric. A symmetrized version, $\mathbb{K}_S$, where

$$\mathbb{K}_S(f_0, f_1) = \mathbb{K}(f_0, f_1) + \mathbb{K}(f_1, f_0)$$

is therefore sometimes used.