On the Tightness of Pinsker's Bound

Wouter M. Koolen

2020-03-04

Abstract

We look into the tightness of Pinsker's bound relating Total Variation distance to Kullback Leibler divergence.

1 Introduction

Let P and Q be probability distributions. Two famous quantities capturing the dissimilarity between P and Q are the Total Variation (TV) distance and the Kullback-Leibler (KL) divergence. Pinsker's inequality relates these measures, stating that

$$\mathrm{TV}(P,Q) \leq \sqrt{\mathrm{KL}(P,Q)/2}.$$

In this post we aim to find the tightest function f for which

$$\mathrm{TV}(P,Q) \leq f(\mathrm{KL}(P,Q)).$$

More specifically, we will analyse

$$f(x) := \max_{P,Q} \ \mathrm{TV}(P,Q) \quad \text{subject to} \quad \mathrm{KL}(P,Q) \leq x. \tag{1}$$

2 Reduction to the Bernoulli case

We first argue that for our problem (1), it suffices to consider distributions on two outcomes, i.e. Bernoulli distributions. To see why, recall that

$$\mathrm{TV}(P,Q) = \sup_{\text{event E}} |P(E) - Q(E)|$$

Now consider some P and Q and let their TV distance be maximised (or nearly so) at some event E. Then if we project P and Q down to Bernoulli $P(E)$ and $Q(E)$, we preserve the TV and decrease the KL (this is the contraction property of the KL).

3 Computation

Restricting to binary candidates, and breaking symmetry arbitrarily, problem (1) reduces to

$$\max_{\substack{0 \le q \le p \le 1 \\ \text{KL}(p,q) \le x}} p - q$$

Using duality, this is

$$\min_{\lambda \ge 0} \max_{0 \le q \le p \le 1} p - q + \lambda(x - \text{KL}(p,q))$$

Reparamterising by $c = p - q \ge 0$ we find

$$\min_{\lambda \ge 0} \max_{0 \le p - c \le p \le 1} c + \lambda(x - \text{KL}(p, p - c))$$

Interestingly, we can solve for c yielding

$$c = \frac{1}{2}\left(-\lambda + \sqrt{\lambda(\lambda - 4p + 2) + 1} + 2p - 1\right)$$

Moreover, we can solve for p, yielding

$$p = \frac{e^{1/\lambda}\left(\left(e^{1/\lambda} - 1\right)\lambda - 1\right)}{\left(e^{1/\lambda} - 1\right)^2 \lambda}$$

resulting in value

$$\inf_{\lambda \ge 0} \frac{1}{e^{1/\lambda} - 1} + \lambda\left(\ln\left(e^{1/\lambda} - 1\right) + \ln\lambda - 1\right) + \lambda x$$

Note that this is a convex optimisation problem (it should be, by the path we took to arrive here, but it can also be verified explicitly using the second derivative). The objective is unbounded below when $x < 0$, as it should be. This function does not seem to analytically simplify further. As it is a one-dimensional convex optimisation problem, we may evaluate it numerically. Here is a plot of the result:
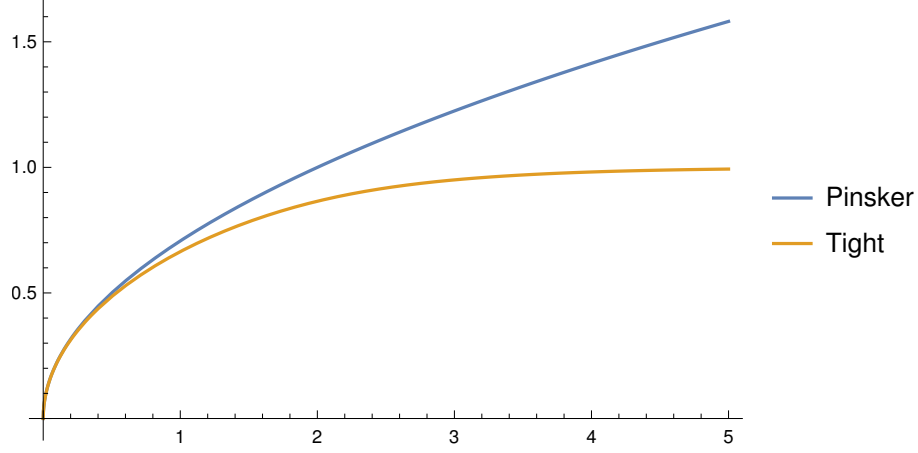
Figure 1: Quality of Pinsker's bound $\sqrt{x/2}$ and the tightest possible bound $f(x)$ from (1)

As $\mathrm{TV}(P, Q) \leq 1$, it makes sense that $f(x)$ increases to 1. This in contrast to Pinsker's bound, which increases to infinity. We see that Pinsker becomes tight near zero Kullback-Leibler divergence.

4   Reparameterisations

Let us conclude with some reparameterisations, which could be helpful. First, we reparameterise by $\eta = 1/\lambda$ to find

$$f(x) = \inf_{\eta \geq 0} \frac{1}{e^\eta - 1} + \frac{\ln\left(e^\eta - 1\right) - \ln \eta - 1 + x}{\eta}$$

Okay, now

$$g(\eta) = \frac{1}{e^\eta - 1} + \frac{\ln\left(e^\eta - 1\right) - \ln \eta - 1}{\eta}$$

is an increasing, concave function. So the answer we are looking for is the function $f(x) = \sup_{\eta \geq 0} g(\eta) + x/\eta$. Bounding $g(\eta)$ from above by $\eta/8$ (the tangent at $\eta = 0$) gives standard Pinsker. To see this, note that

$$f(x) = \inf_{\eta \geq 0} g(\eta) + \frac{x}{\eta} \leq \inf_{\eta \geq 0} \frac{\eta}{8} + \frac{x}{\eta} = \sqrt{x/2}.$$

We may further use the Lambert function to reparameterise by $z = \frac{e^\eta - 1}{\eta}$, resulting in $\eta = -W_{-1}\left(-\frac{e^{-1/z}}{z}\right) - \frac{1}{z}$, to find

$$g(z) = \frac{z - z \ln z - 1}{z W_{-1}\left(-\frac{e^{-1/z}}{z}\right) + 1}$$

3

So we are left with

$$f(x) = \inf_{z \geq 1} \; -\frac{xz - z + z \ln z + 1}{zW_{-1}\left(-\frac{e^{-1/z}}{z}\right) + 1}$$

or even with $y = 1/z$,

$$f(x) = \inf_{y \in [0,1]} \frac{-x - y + \ln y + 1}{W_{-1}\left(-ye^{-y}\right) + y}$$

or even more with $q = -ye^{-y}$,

$$f(x) = \inf_{q \in [-e^{-1},0]} \frac{x - \ln(-q) - 1}{W(q) - W_{-1}(q)}$$