Closed-form formula for the Chernoff information between categorical distributions

Frank Nielsen

November 8, 2024

1 Introduction

We consider discrete probability distributions on a finite alphabet of size $k \geq 2$.

By interpreting the set of categorical distributions as an exponential family, we get the following characterization [?]:

In particular, we get a closed-form solution [?] for binary alphabets (when $k = 2$, the underlying exponential family is of order 1):

2 Characterization and approximation method

Exponential arc.

3 Bernoulli case

For $x \in \mathcal{X} = \{0, 1\}$, the PMF of a Bernoulli distribution is:

$$p^x (1-p)^{1-x}$$

$$D_{\mathrm{KL}}(b_\lambda : b_p) = D(b_\lambda : b_q).$$

Let $a = \frac{p_1}{p_2}$ and $b = \frac{1-p_1}{1-p_2}$. Then we have the optimal Chernoff exponent given by:

$$\alpha^*(p_1, p_2) = \frac{\log\left(\frac{\log b}{\log a}(1 - \frac{1}{p_2})\right)}{\log \frac{a}{b}}.$$

Let $c = \frac{\log(-\frac{\log b}{\log a})}{\log \frac{a}{b}}$. Then the Chernoff distribution is

$$p^* = \frac{1}{1 + (\frac{b}{a})^c}$$

and the Chernoff information is:

$$D_C(p_1, p_2) = D_{B,\alpha^*}(p_1, p_2),$$

where $= D_{B,\alpha^*}$ is the skew Bhattacharyya distance:

$$D_{B,\alpha}(p_1, p_2) = -\log\left(p_1^\alpha p_2^{1-\alpha} + (1-p_1)^\alpha (1-p_2)^{1-\alpha}\right).$$

Example 1 *Consider $p_1 = 0.1$ and $p_2 = 0.2$, we find $\alpha^* \approx 0.4761245029727815$. The Chernoff distribution is $p^* = 0.1452443543242726$. The Chernoff information is about* 0.01012451657995914.

4 Multinoulli case

5 Multiple distributions

natural neighbour interpolation [**?**]

References

[1] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.

[2] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

[3] Frank Nielsen. An information-geometric characterization of Chernoff information. *IEEE Signal Processing Letters*, 20(3):269–272, 2013.

[4] Robin Sibson. A brief description of natural neighbour interpolation. *Interpreting multivariate data*, pages 21–36, 1981.

[5] M Brandon Westover. Asymptotic geometry of multiple hypothesis testing. *IEEE transactions on information theory*, 54(7):3327–3329, 2008.

A Bernoulli exponential family

- $\theta = \log \frac{p}{1-p}$
- $F(\theta) = \log(1 + e^\theta)$
- $\eta(\theta) = \frac{e^\theta}{1+e^\theta}$, $\eta(p) = p$
- $\theta(\eta) = \log \frac{\eta}{1-\eta}$
- $F^*(\eta) = \eta \log \eta + (1-\eta)\log(1-\eta) = -H(b_p)$
- $\lambda(\theta) = \frac{e^\theta}{1+e^\theta}$
- $\lambda_\alpha = \frac{\left(\frac{p_1}{1-p_1}\right)^\alpha \left(\left(\frac{p_2}{1-p_2}\right)^{1-\alpha}\right.}{1+\left(\frac{p_1}{1-p_1}\right)^\alpha \left(\left(\frac{p_2}{1-p_2}\right)^{1-\alpha}\right.}$.

Bernoulli Kullback-Leibler divergence also called binary Kullback-Leibler divergence.
k-ary Kullback-Leibler divergence.