Math. Commun. **18**(2013), 75–78

Un upper bound for Kullback-Leibler divergence with a small number of outliers

ALEXANDER GOFMAN[1] AND MARK KELBERT[2,*]

[1] *Faculty of Economics, Moscow Economics National Research University, 20 Myasnitskaya str., Moscow, 101 000, RF*
[2] *Department of Mathematics, Swansea University, Singleton Park, Swansea, SA2 8PP, UK*

Received November 24, 2011; accepted October 22, 2012

Abstract. We establish a new upper bound for the Kullback-Leibler divergence of two discrete probability distributions which are close in a sense that typically the ratio of probabilities is nearly one and the number of outliers is small.

AMS subject classifications: 94A17, 62B10

Key words: Kullback-Leibler divergence, relative entropy, mutual information, information inequality

---

1. Introduction

Let $\mathbb{P} = (p_i)$ and $\mathbb{Q} = (q_i)$ be two probability distributions on the set $\mathcal{R} = \{1, 2, \dots, r\}$ and

$$D(\mathbb{P} \parallel \mathbb{Q}) = \sum_{i=1}^{r} p_i \ln \frac{p_i}{q_i} \tag{1}$$

the Kullback-Leibler divergence between these distributions (cf. [1, 3]). The problem of obtaining good upper and lower bounds for the relative entropy attracts considerable interest in information theory (see $[1-4]$). We present here two famous upper bounds (see proofs in [2]):

$$D(\mathbb{P} \parallel \mathbb{Q}) \leq \min \Big[\sum_{i=1}^{r} \frac{p_i^2}{q_i} - 1, \sum_{i=1}^{r} \sqrt{\frac{p_i}{q_i}} |p_i - q_i| \Big]. \tag{2}$$

In the case $|\frac{p_i}{q_i} - 1| \leq \epsilon$ for all i the first bound gives an estimate of order $O(\epsilon)$. In the second bound the distributions should be close in the $\mathbf{L}_1$ sense. We present here a new upper bound for $D(\mathbb{P} \parallel \mathbb{Q})$ of the first type. However, we achieve the order of approximation $O(\epsilon^2)$ and also allow a possibility of outliers for $\mathbb{P}$ and $\mathbb{Q}$ with a small total mass. We say that $(\mathbb{P}, \mathbb{Q}) \in \mathcal{C}(\epsilon, \delta, c, d)$ if there exists a partition $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$ where $\mathcal{R}_1$ is called a regular set and $\mathcal{R}_2$ is the set of outliers with the following properties. Assume that for some $\epsilon, \delta > 0$ and some constant $0 < c < 1 < d < \infty$

---

*Corresponding author. *Email addresses:* agofman@rambler.ru (A. Gofman), m.kelbert@swansea.ac.uk (M. Kelbert)

$$\left|\frac{q_i}{p_i}-1\right| \le \epsilon \ \ \forall i \in \mathcal{R}_1, \sum_{j\in\mathcal{R}_2} q_j = \delta, c \le \frac{q_j}{p_j} \le d \ \ \forall j \in \mathcal{R}_2. \tag{3}$$

The main result is the upper bound of the Kullback-Leibler divergence for small enough $\epsilon, \delta > 0$:

Theorem 1. *For any* $(\mathbb{P},\mathbb{Q}) \in \mathcal{C}(\epsilon,\delta,c,d)$

$$0 \le D(\mathbb{P} \parallel \mathbb{Q}) \le \phi(\epsilon) + \psi(\epsilon,\delta), \tag{4}$$

where $\phi(\epsilon) = \frac{1}{2}\epsilon^2 + 0(\epsilon^3)$, $\psi(\epsilon,\delta) \le C\delta$ *for* ϵ, δ *small enough, and* $C = 2d-c-1+\ln\frac{1}{c}$.

Explicit formulas for the functions ϕ and ψ are presented in (12 a,b).

Corollary 1. *Suppose that the joint distribution* $P_{X,Y}$ *on* $\mathcal{R}^{\times 2}$ *of random variables* X *and* Y *satisfies for all* i,j *the inequality*

$$|p_{X,Y}(i,j)/p_X(i)p_Y(j) - 1| \le \epsilon.$$

Then the mutual information between X *and* Y

$$I(X:Y) = D(p_{X,Y} \parallel p_X \times p_Y) \le \phi(\epsilon).$$

2. Bound on the ratio of arithmetic and geometric means

Let $0 < c < a < b < d < \infty$, $x_i \in [a,b], i = 1,\ldots,r_1$, $y_j \in [c,a) \cup (b,d], j = 1,\ldots,r_2$ and $\sum_{j=1}^{r_2} y_j = \delta, r = r_1 + r_2$. Following the strategy used in [5], write $x_i = \lambda_i a + (1-\lambda_i)b$ and $y_j = \mu_j c + (1-\mu_j)d$ with $0 \le \lambda_i, \mu_j \le 1$. Then for any distribution $\mathbb{P}$ on $\mathcal{R}$ and any convex function f on $[c,d]$

$$0 \le \sum_{i=1}^{r_1+r_2} p_i f(t_i) - f(\sum_{i=1}^{r_1+r_2} p_i t_i) = S - T \tag{5}$$

with $t_i \in \{x_i\} \cup \{y_j\}$. Let us estimate the first sum as follows

$$\begin{aligned} S &\le \sum_{i=1}^{r_1} p_i\big(\lambda_i f(a) + (1-\lambda_i) f(b)\big) + \sum_{i=1}^{r_2} p_i\big(\mu_i f(c) + (1-\mu_i) f(d)\big) \\ &= (f(a)-f(b))p_\lambda + f(b)(1-\delta) + (f(c)-f(d))p_\mu + f(d)\delta \end{aligned}$$

with $p_\lambda = \sum_{i=1}^{r_1} p_i \lambda_i, p_\mu = \sum_{j=1}^{r_2} p_j \mu_j \le \delta$.

If we select the convex function as $f(x) = -\ln x$ inequality (5) may be considered as a logarithmic form of the inequality for the ratio of arithmetic and geometric means. We proceed by specifying the second term

$$-T = \ln\big((a-b)p_\lambda + b(1-\delta) + (c-d)p_\mu + d\delta\big)$$

and using the Taylor expansion

$$-T = \ln\big((a-b)p_\lambda + b(1-\delta) + d\delta\big) + \frac{1}{\xi}(c-d)p_\mu$$

with $(a-b)p_\lambda + b(1-\delta) + d\delta + (c-d)p_\mu < \xi < (a-b)p_\lambda + b(1-\delta) + d\delta$. So, the RHS of (5) for the function $f(x) = -\ln x$ is estimated as follows:

$$S - T \le (\ln\frac{b}{a})p_\lambda + (\ln\frac{d}{c})p_\mu + C_1 + \ln\big[(a-b)p_\lambda + b(1-\delta) + d\delta\big] - \frac{d-c}{\xi}p_\mu \quad (6)$$

with $C_1 = -(1-\delta)\ln b - \delta\ln d$. An upper bound will be obtained if we estimate ξ from above $\xi < b(1-\delta) + d\delta$ as $c < d$. This result provides an upper bound which is a linear function of p_μ:

$$S - T \le \big(\ln\frac{b}{a}\big)p_\lambda + K(d,c,b,\delta)p_\mu + C_1 + \ln\big[(a-b)p_\lambda + b(1-\delta) + d\delta\big]. \quad (7)$$

In case the coefficient

$$K = K(d,c,b,\delta) = \ln\frac{d}{c} - \frac{d-c}{b(1-\delta)+d\delta} > 0,$$

an upper bound is obtained by selecting $p_\mu = \delta$. In the case $K < 0$ we select $p_\mu = 0$. In any case, the term proportional to p_μ does not depend on p_λ and the value p^*_λ maximizing the RHS in (7) is the solution of equation

$$\ln b - \ln a + \frac{a-b}{(a-b)p_\lambda + b(1-\delta) + d\delta} = 0, \ \text{ i.e. } p^*_\lambda = \frac{b(1-\delta)+d\delta}{b-a} - \frac{1}{\ln b - \ln a}. \quad (8)$$

It is straightforward to check that the second derivative is negative at p^*_λ. Next, note that

$$p^*_\lambda > \frac{b}{b-a} - \frac{1}{\ln b - \ln a} > 0$$

and $p^*_\lambda < 1-\delta$ for $\delta < \delta^* = \frac{1}{d-a}\big(\frac{b-a}{\ln b - \ln a} - a\big)$.

Let $L(a,b) = \frac{b-a}{\ln b - \ln a}$. Combining all results together we obtain that

$$S - T \le \ln\Lambda(a,b) + \delta\Gamma(a,b,c,d,\delta), \quad (9)$$

where

$$\ln\Lambda(a,b) = \frac{b}{L(a,b)} - 1 - \ln b + \ln(L(a,b)). \quad (10)$$

The function $\Lambda(a,b)$ in first term may be represented as follows (cf. [5]):

$$\Lambda(a,b) = abL(a,b)/G^2(a,b), G(a,b) = \frac{1}{e}\big(b^b/a^a\big)^{1/(b-a)}.$$

in the case $K(d,c,b,\delta) > 0$ we have

$$\Gamma(a,b,c,d,\delta) = \frac{(d-a)\ln b - (d-b)\ln a}{b-a} - \ln c + \frac{d-c}{b+(d-b)\delta} \quad (11a)$$

and in the case $K(d,c,b,\delta) \le 0$

$$\Gamma(a,b,c,d,\delta) = \frac{(d-a)\ln b - (d-b)\ln a}{b-a} - \ln c. \quad (11b)$$

Observe that $\Gamma(a,b,c,d,\delta) > 0$ in both cases.

3. Proof of the Theorem 1

Let us apply the results of the previous section with $t_i = \frac{q_i}{p_i}, a = 1 - \epsilon, b = 1 + \epsilon$. By straightforward calculations the expression in (10) admits the expansion

$$\ln \Lambda(1-\epsilon, 1+\epsilon) = \epsilon + \frac{1}{3}\epsilon^2 - \epsilon + \frac{1}{2}\epsilon^2 - \frac{1}{3}\epsilon^2 + \ldots = \frac{1}{2}\epsilon^2 + \ldots$$

For ϵ small enough both expressions in (11a,b) are bounded by $2d - c - 1 + \ln \frac{1}{c}$ by inspection. Now the Theorem follows from (9) with

$$\phi(\epsilon) = (1+\epsilon)F(\epsilon)^{-1} - 1 - \ln(1+\epsilon) - \ln F(\epsilon), \quad F(\epsilon) = \Big(\ln \frac{1+\epsilon}{1-\epsilon}\Big)/2\epsilon \qquad (12a)$$

and

$$\psi(\epsilon, \delta) = \delta\Big[(d - 1 + \epsilon)F(\epsilon) - \ln c + \frac{d-c}{1 + (d-1-\epsilon)\delta}\Big]. \qquad (12b)$$

References

[1] T. M. Cover, J. M. Thomas, *Elements of Information Theory*, Wiley, New York, 2006.

[2] S. S. Dragomir, V. Gluščević, *New estimates of the Kullback-Leibler distance and applications*, in: *Inequality Theory and Applications Vol. I*, (Y. J. Cho, J. K. Kim and S. S. Dragomir, Eds.), Nova Sci. Publ., Huntington, NY, 2001, 123–137.

[3] M. Kelbert, Y. Suhov, *Information Theory and Coding by Example*, Cambridge Univ. Press, Cambridge, 2013.

[4] A. Sayyareh, *A new upper bound for Kullback-Leibler divergence*, Appl. Math. Sci. **67**(2011), 3303–3317.

[5] S. Simic, *Best possible global bounds for Jensen functional*, AMS Proceedings **138**(2010), 2457–2462.