Guessing Under Source Uncertainty With Side Information

Rajesh Sundaresan, *Senior Member, IEEE*

***Abstract*— We study the problem of guessing the realization of a finite alphabet source, when some side information is provided, in a setting where the only knowledge the guesser has about the source and the correlated side information is that the joint source is one among a family. We define a notion of redundancy, identify a quantity that measures this redundancy, and study its properties. We then identify good guessing strategies that minimize the supremum redundancy (over the family). The minimum value measures the richness of the uncertainty class.**

***Index Terms*— f-divergence, guessing, I-projection, mismatch, Pythagorean identity, redundancy, Rényi information divergence.**

I. Inaccuracy and redundancy in guessing

The problem of guessing has been studied by Massey [1], Arikan [2], and others. In this paper, we study the problem of guessing the realization of a finite alphabet source, when some side information is provided, in a setting where the only knowledge the guesser has about the source and the correlated side information is that the joint source is one among a family. There are several parallels between guessing and source coding ([3], [4]), under source uncertainty. The results in this paper bring these similarities into light.

Let $\mathbb{X}$ and $\mathbb{Y}$ be finite alphabet sets. Consider a correlated pair of random variables (X,Y) with joint PMF P on $\mathbb{X}\times\mathbb{Y}$. Given side information $Y=y$, we would like to guess the realization of X. Formally, a guessing list G with side information is a function $G:\mathbb{X}\times\mathbb{Y}\to\{1,2,\cdots,|\mathbb{X}|\}$ such that for each $y\in\mathbb{Y}$, the function $G(\cdot,y):\mathbb{X}\to\{1,2,\cdots,|\mathbb{X}|\}$ is bijective, and denotes the order in which the elements of $\mathbb{X}$ will be guessed when the guesser observes $Y=y$.

Knowing the PMF P, the best strategy that minimizes the expected number of guesses, given Y, is to guess in the decreasing order of $P(\cdot,Y)$-probabilities. Let us denote such an order G_P. Arikan [2] showed the following general result that gave an operational meaning to the conditional Rényi entropy $H_\alpha(P)$ of order α.

Theorem 1: (Arikan's Guessing Theorem) Let $\rho>0$ and $\alpha=\frac{1}{1+\rho}$. Consider a source pair (X,Y) with PMF P. Then

$$\begin{aligned} H_\alpha(P)-\log(1+\ln|\mathbb{X}|) &\leq \frac{1}{\rho}\log\left(\min_G \mathbb{E}\left[G(X,Y)^\rho\right]\right)\\ &\leq H_\alpha(P).\end{aligned}$$

□

R. Sundaresan is with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560 012, India

This work was supported by the Ministry of Human Resources and Development (MHRD, India) under Grant Part(2A) Tenth Plan (338/ECE).

The conditional Rényi entropy is given by

$$H_\alpha(P)=\frac{\alpha}{1-\alpha}\log\left(\sum_{y\in\mathbb{Y}}\left(\sum_{x\in\mathbb{X}}P(x,y)^\alpha\right)^{1/\alpha}\right).$$

Due to lack of exact knowledge of P, suppose we guess using some guessing list G. Associated with G is a PMF Q_G such that G proceeds in the decreasing order of Q_G-probabilities. If Q_G and P do not lead to the same order, *i.e.*, $G\neq G_P$, then the guessing is *mismatched*. Let us define the *redundancy* in guessing X with side information Y, when the source is P, as follows:

$$R(P,G)\triangleq\frac{1}{\rho}\log\left(\mathbb{E}\left[G(X,Y)^\rho\right]\right)-\frac{1}{\rho}\log\left(\mathbb{E}\left[G_P(X,Y)^\rho\right]\right) \tag{1}$$

The expectation $\mathbb{E}$ in (1) is with respect to P. The dependence of $R(P,G)$ on ρ is understood and suppressed. The following proposition bounds the redundancy on either side.

Theorem 2: Let $\rho>0$, $\alpha=1/(1+\rho)$. Consider a source pair (X,Y) with PMF P. Let G be an arbitrary guessing list with side information Y and Q_G the associated PMF. Then

$$|R(P,G)-L_\alpha(P,Q_G)|\leq\log(1+\ln|\mathbb{X}|).$$

□

Thus the *penalty* in the guessing moment suffered as a result of the mismatch is given by the quantity $L_\alpha(P,Q_G)$, where

$$\begin{aligned}&L_\alpha(P,Q)\triangleq\\ &\quad\frac{\alpha}{1-\alpha}\log\left(\sum_{y\in\mathbb{Y}}\sum_{x\in\mathbb{X}}P(x,y)\left[\sum_{a\in\mathbb{X}}\left(\frac{Q(a,y)}{Q(x,y)}\right)^\alpha\right]^{\frac{1-\alpha}{\alpha}}\right)\\ &\quad-\ H_\alpha(P),\end{aligned} \tag{2}$$

to within $\log(1+\ln|\mathbb{X}|)$. The expression for L_α in (2) specializes to the known expression [5] for the case without side information when $|\mathbb{Y}|=1$.

For this special case, a universal guessing strategy that guesses in the increasing order of empirical entropy was proposed by Arikan and Merhav in [6]. Their strategy is universal inasmuch as it is asymptotically optimal (within $O((\log n)/n)$) for all finite-alphabet, memoryless sources. In our work, we are interested in understanding and interpreting this universality. Moreover, we would like to study and identify good guessing strategies that work well over richer classes of sources, and also generalize to the case with side information.

The quantity $L_\alpha(P,Q)$ also arises in the context of redundancy for Campbell's average exponential coding length problem with side information. (The results for the case with

side information are simple generalizations of [7], [8]). In that case, the values that the parameter ρ takes are expanded to $-1 < \rho < 0$ (resp. $1 < \alpha < \infty$) and $0 < \rho < \infty$ (resp. $0 < \alpha < 1$). Our results below on the properties of $L_\alpha(P, Q)$ are valid for all these α's.

For the case when there is no side information, *i.e.*, $|\mathbb{Y}| = 1$, it is known that $L_\alpha(P, Q)$ can be written in terms of some well-studied divergence quantities. Indeed,

$$L_\alpha(P, Q) = D_\beta(P'||Q') = \frac{1}{\rho} \log \left(\text{sign}(\rho)\ I_f(P'||Q') \right), \quad (3)$$

where $\beta = 1/\alpha = 1 + \rho$, $D_\beta(R||S)$ is the Rényi divergence of order β (see for example [9]), I_f is Csiszár's f-divergence (see [10]) with $f(x) = \text{sign}(\rho) \cdot x^{1+\rho}$, and P' is the tilted PMF obtained from P and given by

$$P'(x) = \frac{P(x)^\alpha}{\sum_{a \in \mathbb{X}} P(a)^\alpha}.$$

It is known that $\lim_{\alpha \to 1} L_\alpha(P, Q) = D(P||Q)$, the Kullback-Leibler divergence, for the special case. For the case with side information, $\lim_{\alpha \to 1} L_\alpha(P, Q) = D(P_{X|Y} \parallel Q_{X|Y} \mid P_Y)$, where (X, Y) is the pair of random variables of interest with PMF $P = P_Y \cdot P_{X|Y}$. The mismatched PMF $Q = Q_Y \cdot Q_{X|Y}$.

II. Problem Statement

Let $\mathbb{T}$ denote a set of PMFs on the finite alphabet $\mathbb{X} \times \mathbb{Y}$. $\mathbb{T}$ may be infinite in size. Associated with $\mathbb{T}$ is a family $\mathcal{T}$ of measurable subsets of $\mathbb{T}$ and thus $(\mathbb{T}, \mathcal{T})$ is a measurable space. We assume that for every $x \in \mathbb{X}$, the mapping $P \mapsto P(x)$ is $\mathcal{T}$-measurable.

For a fixed $\rho > 0$, we seek a good guessing strategy G that works well for all $P \in \mathbb{T}$. G can depend on knowledge of $\mathbb{T}$, but not on the actual source PMF. More precisely, for $P \in \mathbb{T}$ the redundancy denoted by $R(P, G)$ when the true source is P and when the guessing list is G, is given by (1). The worst redundancy under this guessing strategy is given by

$$\sup_{P \in \mathbb{T}} R(P, G)$$

Our first aim is to minimize this worst redundancy over all guessing strategies, *i.e.*, find a G that attains the minimum in

$$\min_G \sup_{P \in \mathbb{T}} R(P, G)$$

In view of Theorem 2, clearly, the following quantity is relevant for $0 < \alpha < 1$. The definition however is wider in scope.

Definition 3: For $\alpha > 0, \alpha \neq 1$,

$$C \stackrel{\Delta}{=} \min_Q \sup_{P \in \mathbb{T}} L_\alpha(P, Q). \quad (4)$$

One contribution of this paper is to demonstrate the existence of a minimizing Q^* thereby justifying the use of "min" instead of "inf". For the case when $|\mathbb{T}|$ is finite and $|\mathbb{Y}| = 1$, previously known results for f-divergences and Rényi divergences can be used to show the existence and characterization of Q^*. Here, however, $\mathbb{T}$ may have an infinite number of elements or $|\mathbb{Y}| > 1$ or both.

Theorem 4: There exists a unique PMF Q^* such that

$$C = \sup_{P \in \mathbb{T}} L_\alpha(P, Q^*) = \inf_Q \sup_{P \in \mathbb{T}} L_\alpha(P, Q).$$

□

The minimizing Q^* has the geometric interpretation of a *center* of the uncertainty set $\mathbb{T}$. Accordingly, C plays the role of *radius*; all elements in the uncertainty set $\mathbb{T}$ are within a "squared distance" C from the center Q^*. The reason for describing $L_\alpha(P, Q)$ as "squared distance" is because it shares the Pythagorean property with Euclidean squared distance and with Kullback-Leibler divergence. See the next section for more details on this other contribution of the paper. It is already known that L_α satisfies this property in the no side information case [5]; our results in the next section show that the property holds in a wider context.

Returning to the problem of guessing, the following result shows that guessing in the decreasing order of Q^*-probabilities, where Q^* attains the min-sup in Definition 3, results in min-sup redundancy to within $\log(1 + \ln |\mathbb{X}|)$.

Theorem 5: (*Guessing under uncertainty*) Consider the class of PMFs parameterized by $\mathbb{T}$. There exists a guessing list G^* for X with side information Y such that

$$\sup_{P \in \mathbb{T}} R(P, G^*) \leq C + \log(1 + \ln |\mathbb{X}|).$$

In particular, $G^* = G_{Q^*}$. Conversely, for any arbitrary guessing strategy G, the worst-case redundancy is at least $C - \log(1 + \ln |\mathbb{X}|)$, *i.e.*,

$$\sup_{P \in \mathbb{T}} R(P, G) \geq C - \log(1 + \ln |\mathbb{X}|).$$

□

The converse part of Theorem 5 is meaningful only when $C > \log(1 + \ln |\mathbb{X}|)$. This will hold, for example, when the uncertainty class is sufficiently rich. The finite state, arbitrarily varying source is one such example. Observe that if we have $\mathbb{X} \times \mathbb{Y} = \mathbb{A}^n \times \mathbb{B}^n$, a cartesian product of the n-fold cartesian product of $\mathbb{A}$ and $\mathbb{B}$, then $\log(1 + \ln |\mathbb{X}|)$ grows logarithmically with n if $|\mathbb{X}| \geq 2$. The uncertainty class is therefore rich enough for the converse to be meaningful if C grows with n at a faster rate.

This brings us to the interpretation of the universality result in [6] where Arikan and Merhav proposed a guessing strategy that achieves the guessing exponent to within $O((\log n)/n)$ for all finite-alphabet memoryless sources (without side information). We interpret this as follows: the normalized L_α-radius for the class of finite-alphabet memoryless sources vanishes asymptotically as $O((\log n)/n)$, a fact that we demonstrate rather directly in [11] by choosing a candidate for center and enclosing all PMFs in the class within normalized radius $O((\log n)/n)$ from the chosen center.

The above results are analogous to results on source coding under source uncertainty [3], [4]. The channel capacity of an associated channel plays the role of L_α-radius in this paper. The geometric results of the next section shed some light on the nature of Q^* and may be of independent interest.

III. L_α-PROJECTION : A GENERALIZATION

In this section, we generalize the Pythagorean property to L_α with side information. The results of this section are relevant to the guessing problem to the extent that they characterize the minimizing Q^*, and are straightforward extensions of results in [12] and [5]. Note that the expression for L_α is different from [5] and accounts for side information. We proceed along the same lines.

Let $\mathbb{X}$ and $\mathbb{Y}$ be finite alphabet sets. Given a PMF R on $\mathbb{X} \times \mathbb{Y}$, the set of PMFs on $\mathbb{X} \times \mathbb{Y}$

$$B(R, r) \stackrel{\Delta}{=} \{P \mid L_\alpha(P, R) < r\}, \quad 0 < r \leq \infty,$$

is called an L_α-sphere (or ball) with center R and radius r. The term "sphere" conjures the image of a convex set. That the set is indeed convex needs a proof since $L_\alpha(P, R)$ is not convex in its arguments.

Proposition 6: $B(R, r)$ is a convex set. □

When we talk of closed sets, we refer to the usual Euclidean metric on the $|\mathbb{X}||\mathbb{Y}|$-dimensional Euclidean vector space. The set of PMFs on $\mathbb{X} \times \mathbb{Y}$ is closed and bounded (and therefore compact).

If $\mathcal{E}$ is a closed and convex set of PMFs on $\mathbb{X}\times\mathbb{Y}$ intersecting $B(R, \infty)$, *i.e.* there exists a PMF $P \in \mathcal{E}$ such that $L_\alpha(P, R) < \infty$, then a PMF $Q \in \mathcal{E}$ satisfying

$$L_\alpha(Q, R) = \min_{P \in \mathcal{E}} L_\alpha(P, R),$$

is called the *L_α-projection* of R on $\mathcal{E}$.

Proposition 7: (Existence of L_α-projection) Let $\mathcal{E}$ be a closed and convex set of PMFs on $\mathbb{X} \times \mathbb{Y}$. If $B(R, \infty) \cap \mathcal{E}$ is nonempty, then R has an L_α-projection on $\mathcal{E}$. □

We next state the generalizations of [12, Lemma 2.1, Theorem 2.2]. Here $L_\alpha(P, Q)$ plays the role of squared Euclidean distance (analogous to the Kullback-Leibler divergence).

Proposition 8: Let $0 < \alpha < \infty, \alpha \neq 1$.

1) Let $L_\alpha(Q, R)$ and $L_\alpha(P, R)$ be finite. The segment joining P and Q does not intersect the L_α-sphere $B(R, r)$ with radius $r = L_\alpha(Q, R)$, *i.e.*,

$$L_\alpha(P_\lambda, R) \geq L_\alpha(Q, R)$$

for each

$$P_\lambda = \lambda P + (1 - \lambda)Q, \quad 0 \leq \lambda \leq 1,$$

if and only if

$$L_\alpha(P, R) \geq L_\alpha(P, Q) + L_\alpha(Q, R). \tag{5}$$

2) *(Tangent hyperplane)* Let

$$Q = \lambda P + (1 - \lambda)S, \quad 0 < \lambda < 1. \tag{6}$$

Let $L_\alpha(Q, R)$, $L_\alpha(P, R)$, and $L_\alpha(S, R)$ be finite. The segment joining P and S does not intersect $B(R, r)$ (with $r = L_\alpha(Q, R)$) if and only if

$$L_\alpha(P, R) = L_\alpha(P, Q) + L_\alpha(Q, R). \tag{7}$$

□

Proposition 8.2 extends the analog of Pythagoras theorem, known to hold for the Kullback-Leibler divergence, and for L_α without side information. Let us now apply Proposition 8 to the L_α-projection of a convex set.

For a convex $\mathcal{E}$, we call Q an algebraic inner point of $\mathcal{E}$ if for every $P \in \mathcal{E}$, there exist λ and S satisfying (6).

Theorem 9: (Projection Theorem) Let $0 < \alpha < \infty, \alpha \neq 1$. A joint PMF $Q \in \mathcal{E} \cap B(R, \infty)$ is the L_α-projection of R on the convex set $\mathcal{E}$ if and only if every $P \in \mathcal{E}$ satisfies

$$L_\alpha(P, R) \geq L_\alpha(P, Q) + L_\alpha(Q, R). \tag{8}$$

If the L_α-projection Q is an algebraic inner point of $\mathcal{E}$, then every $P \in \mathcal{E} \cap B(R, \infty)$ satisfies (8) with equality. □

While existence of L_α-projection is guaranteed for certain sets by Proposition 7, we can show that a projection onto a convex set, if it exists, is unique.

As an application of Theorem 9, let us characterize the L_α-center of a family.

Proposition 10: The L_α-center of $\mathbb{T}$ lies in the closure of its convex hull. □

The proof is quite simple. If the L_α-center does not lie in the closure of the convex hull of $\mathbb{T}$, the projection of the PMF on the closed and convex set is a better candidate.

IV. SUMMARY OF MAIN CONTRIBUTIONS

The main contributions of this paper are as follows:

- a highlighting of the similarity between guessing and source coding;
- a generalization of the results on redundancy in guessing to the case when side information is available;
- a discovery of the fact that the relevant quantity that measures redundancy in guessing with side information also satisfies the Pythagorean property.

All the above are generalizations of known results for the case without side information. Proofs of stated results are available in [11].

REFERENCES

[1] J. L. Massey, "Guessing and entropy," in *Proc. 1994 IEEE International Symposium on Information Theory*, Trondheim, Norway, Jun. 1994, p. 204.

[2] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 99–105, Jan. 1996.

[3] L. D. Davisson, "Universal noiseless coding," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 783–795, Nov. 1972.

[4] R. G. Gallager, "Source coding with side information and universal coding," *LIDS Technical Report, LIDS-P-937*, Sept. 1979.

[5] R. Sundaresan, "A measure of discrimination and its geometric properties," in *Proc. 2002 IEEE Int. Symp. on Information Theory*, Lausanne, Switzerland, Jun. 2002, p. 264.

[6] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 1041–1056, May 1998.

[7] L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, pp. 423–429, 1965.

[8] ——, "Definition of entropy by means of a coding problem," *Z. Wahrscheinlichkeitstheorie verw. Geb.*, vol. 6, pp. 113–118, 1966.

[9] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Trans. Inform. Theory*, vol. IT-41, pp. 26–34, Jan. 1995.

[10] ——, "A class of measures of informativity of observation channels," *Periodica Mathematica Hungarica*, vol. 2, pp. 191–213, 1972.

[11] R. Sundaresan, "Guessing under source uncertainty," March 2006, submitted to *IEEE Trans. Inform. Theory*, available at http://www.arxiv.org/abs/cs.IT/0603064.

[12] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *The Annals of Probability*, vol. 3, pp. 146–158, 1975.