On Guessing The Realization Of An Arbitrarily Varying Source

Rajesh Sundaresan
Department of Electrical Communication
Engineering
Indian Institute of Science
Bangalore 560012
rajeshs@ece.iisc.ernet.in

Abstract — **We present a method to guess the realization of an arbitrarily varying source. Let T_U be the type of the unknown state sequence. Our method results in a guessing moment that is within $K_n(T_U) + O(\log n/n)$ of the minimum attainable guessing moment with full knowledge of source statistics, *i.e.*, with knowledge of the sequence of states s^n. The quantity $K_n(T_U) + O(\log n/n)$ can be interpreted as the penalty one pays for not knowing the sequence of states s^n of the source. $K_n(T_U)$ by itself is the penalty one pays for guessing with the additional knowledge that the state sequence belongs to type T_U. Conversely, given any guessing strategy, for every type T_U, there is a state sequence belonging to this type whose corresponding source forces a guessing moment penalty of at least $K_n(T_U) - O(\log n/n)$.**

I. Guessing under source mismatch

Let X be a random variable on a finite set $\mathbb{X}$ with probability mass function (PMF) given by $(P(x) : x \in \mathbb{X})$. Consider the problem of guessing the realization of this random variable X by asking questions of the form “Is X equal to x?”, stepping through the elements of $\mathbb{X}$, until the answer is “Yes” ([1], [2]).

Massey [1] and Arikan [2] considered guessing strategies, *i.e.*, sequences of guesses, and sought to lowerbound the minimum expected number of guesses. For a given guessing strategy G, let $G(x)$ denote the number of guesses required when $X = x$. The strategy that minimizes the expected number of guesses, $E[G(X)]$, proceeds in the decreasing order of source probabilities. Let us denote this optimum guessing order that depends on the source PMF P by G_P. Arikan [2] showed that the exponent of the minimum value, *i.e.*, $\log[\min_G \; E[G(X)]] = \log[E[G_P(X)]]$, satisfies

$$H_{1/2}(P) - \log(1 + \ln|\mathbb{X}|) \le \log[E[G_P(X)]] \le H_{1/2}(P),$$

where $H_\alpha(P)$ is the Rényi entropy of order $\alpha > 0$.

For $\rho > 0$, Arikan [2] also considered minimization of $(E[G(X)^\rho])^{1/\rho}$ over all guessing strategies G; G_P minimizes this value, and the exponent of the minimum value satisfies [2]

$$H_\alpha(P) - \log(1 + \ln|\mathbb{X}|) \le \frac{1}{\rho}\log[E[G_P(X)^\rho]] \le H_\alpha(P), \tag{1}$$

where $\alpha = 1/(1+\rho)$. Throughout this paper, $\rho > 0$, $\alpha = 1/(1+\rho)$, and therefore $\alpha \in (0,1)$.

Suppose now that we do not know the true PMF P, but guessed assuming a PMF Q, where $Q \neq P$. Let this lead to a guessing strategy G_Q. Thus G_Q is not matched to the source. We might therefore anticipate that $\frac{1}{\rho}\log[E[G_Q(X)^\rho]]$ is larger. Analogously, for any arbitrary guessing strategy G, we may think of an associated PMF Q_G for which G is the optimum guessing strategy. If $Q_G \neq P$, G may not be matched to the source. Setting $\alpha = 1/(1+\rho)$, we claim (without going into details) that for any guessing strategy G, the *redundancy* defined by

$$R(G) \triangleq \frac{1}{\rho}\log[E[G(X)^\rho]] - \frac{1}{\rho}\log[E[G_P(X)^\rho]]$$

satisfies

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

where Q_G is a particular distribution obtained from G.

While the above is true only for some specific PMFs, we actually have the following for any PMF Q:

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

The quantity $L_\alpha(P,Q)$ is given by

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

is therefore a measure of the redundancy (to within $\log(1 + \ln|\mathbb{X}|)$) of the logarithm of the ρth guessing moment, when the *mismatched* PMF Q is used to obtain the guessing strategy. It can be shown that $L_\alpha(P,Q) \ge 0$, and equality is achieved if and only if $P = Q$ [7].

To understand the nature of $L_\alpha(P,Q)$, let

$$P'(\cdot) \triangleq \frac{P(\cdot)^\alpha}{\sum_{a\in\mathbb{X}} P(a)^\alpha}, \quad Q'(\cdot) \triangleq \frac{Q(\cdot)^\alpha}{\sum_{a\in\mathbb{X}} Q(a)^\alpha}. \tag{3}$$

In these definitions, the dependence of the primed PMFs on α is understood and suppressed. $P \mapsto P'$ and $Q \mapsto Q'$ are one-to-one mappings.

Straightforward algebra results in

$$L_\alpha(P,Q) = \frac{1}{\rho}\log\left[I_f(P'||Q')\right], \tag{4}$$

where $I_f(R||S)$ is the f-divergence with $f(x) = x^{1+\rho}$ (see for e.g., [4]) given by

$$I_f(R||S) = \sum_{x\in\mathbb{X}} S(x) f\left(\frac{R(x)}{S(x)}\right). \tag{5}$$

Furthermore, we have $L_\alpha(P,Q) = \infty$ if and only if Support$(P) \not\subset$ Support(Q). (Note that $\alpha \in (0,1)$).

We emphasize that $L_\alpha(P,Q)$ is not a convex function of P or of Q. Moreover, unlike f-divergences, $L_\alpha(P,Q)$ does not satisfy the data processing inequality. However, $L_\alpha(P,Q)$ does satisfy a Pythagorean-type inequality and behaves like "squared-distance". This property was explored in detail in [7].

II. A FINITE FAMILY OF SOURCES

Suppose now that $(P_\theta : \theta \in \Theta)$ is a one-parameter family of sources on the finite alphabet set $\mathbb{X}$. The family is of finite size, *i.e.*, $|\Theta| < \infty$. (In recent work, we have been able to extend some of the results of this section to families of infinite sizes with some mild regularity conditions on the family [8]. This however is beyond the scope of this paper). The true source is one from the family, but is otherwise unknown. A guessing strategy for the above family of sources can be devised as follows. Let G_θ be the guessing order when P_θ is known to be the source. (Note that G_θ is short form for the more cumbersome G_{P_θ}). Now merge these $|\Theta|$ guessing lists as follows:

- First order the elements in Θ and call them $\theta_1, \cdots, \theta_{|\Theta|}$.
- Set $i = 1$.
- List elements of $\mathbb{X}$ in the following order: ith element of G_{θ_1}, then the ith element of G_{θ_2}, and so on, skipping an element if it is already in the list, until the ith element of $G_{\theta_{|\Theta|}}$ is reached.
- Once done with all ith elements, increment i, and continue, until all elements are listed.

An important property of this list is that $G(x) \le |\Theta| G_\theta(x)$, for every $x \in \mathbb{X}$ and for every $\theta \in \Theta$. Note that this is a strategy that does not depend on the parameter θ. We therefore have the following result. Define the penalty function for an arbitrary guessing function G as follows:

$$R(\theta, G) \triangleq \frac{1}{\rho}\log\left[E_\theta\left[G(X)^\rho\right]\right] - \frac{1}{\rho}\log\left[E_\theta\left[G_\theta(X)^\rho\right]\right],$$

where E_θ is expectation with respect to P_θ.

Proposition 1 *For the finite family $(P_\theta : \theta \in \Theta)$, the guessing strategy G obtained by merging the individual guessing lists suffers a maximum penalty which is upperbounded by*

$$\max_{\theta\in\Theta} R(\theta, G) \le \log|\Theta| + \log(1 + \ln|\mathbb{X}|).$$

Now suppose that we wish to do better for this family of sources. One approach is to find a Q^* that comes close to

$$K = \inf_Q \max_{\theta\in\Theta} L_\alpha(P_\theta, Q), \tag{6}$$

where the infimum is over all PMFs on $\mathbb{X}$. We can think of K in (6) as playing the geometric role of minimum *radius* enclosing all points P_θ in the family, measured from the *center* Q^*. This minimum radius is termed the "L_α-radius" of the family $(P_\theta : \theta \in \Theta)$, because "squared-distances" are given by $L_\alpha(\cdot,\cdot)$. We caution that $L_\alpha(\cdot,\cdot)$ is not a *distance* function or a *metric* in the strict sense. Indeed, it is not symmetric in the two arguments.

From (4) and the monotone increasing property of $\log(\cdot)$, the problem in (6) is straightforwardly transformed into the problem of finding the "I_f-radius" and "I_f-center" of the family $(P'_\theta : \theta \in \Theta)$, *i.e.*,

$$K' = \inf_{Q'} \max_{\theta\in\Theta} I_f(P'_\theta, Q'). \tag{7}$$

See [4] for a solution to the problem, and also [5] and [6] for related problems. The problem transformation is done by using the mapping $P_\theta \mapsto P'_\theta$ and the relationship in (4).

Results of [5], [4] and[6] show that the "inf" in (6) can be replaced by a "min", and therefore

$$K' = \min_{Q'} \max_{\theta\in\Theta} I_f(P'_\theta, Q'). \tag{8}$$

The minimizing Q^* can be found from the inverse mapping of $Q^* \mapsto Q'^*$ where Q'^* is the "I_f-center" of the family $(P'_\theta : \theta \in \Theta)$. Furthermore, [7] indicates that this minimizing Q^* is a mixture of the sources that make up the family, *i.e.*,

$$Q^* = \sum_{\theta\in\Theta} w(\theta) P_\theta.$$

This property is a consequence of the Pythagorean-type inequality satisfied by $L_\alpha(P,Q)$. It can also be seen from the results in the following section. Thus the "inf" in 6 can also be replaced by "min".

Returning to the problem of finding a guessing strategy for the family of sources, we can guess in the decreasing order of probabilities given by the PMF Q^*, which then leads to the following result.

Proposition 2 *For the finite family $(P_\theta : \theta \in \Theta)$, let Q^* be the L_α-center, and K the L_α-radius. Let G_{Q^*} be the guessing strategy obtained by guessing in the decreasing order of* *probabilities of Q^*. Then the maximum penalty suffered by the strategy is upperbounded by the following:*

$$\max_{\theta\in\Theta} R(\theta, G_{Q^*}) \leq K + \log\left(1 + \ln|\mathbb{X}|\right).$$

Furthermore, for any guessing strategy G, we have

$$\max_{\theta\in\Theta} R(\theta, G) \geq K - \log\left(1 + \ln|\mathbb{X}|\right).$$

III. Necessary and sufficient conditions for finding the center and radius

The needed results from [4] for finding I_f-center and I_f-radius are first reproduced. These can then be easily transformed to find the L_α-center and L_α-radius.

Csiszar [4] shows that

$$K' = \min_{Q'} \max_{\theta\in\Theta} I_f(P'_\theta, Q') = \max_{v} \min_{Q'} \sum_{\theta\in\Theta} v(\theta) I_f(P'_\theta, Q'). \tag{9}$$

Furthermore, for any given v, the minimizing Q'^* is given by

$$Q'^*(x) = C\left(\sum_{\theta\in\Theta} v(\theta)\left[P'_\theta(x)\right]^{1/\alpha}\right)^\alpha, \; \forall x \in \mathbb{X}. \tag{10}$$

C is a normalizing constant. Also, v^* attains the maximum in (9) if and only if

$$I_f(P'_\theta, Q'^*) \leq K', \; \forall\theta \in \Theta, \tag{11}$$

where equality holds whenever $v(\theta) > 0$.

We make one remark and close this section. By using the inverse of the mapping $P \mapsto P'$ on the minimizing Q'^*, we get that

$$Q^* = \sum_{\theta\in\Theta} w(\theta) P_\theta,$$

where

$$w(\theta) = A\frac{v(\theta)}{\left[\sum_{a\in\mathbb{X}} P_\theta(a)^{1/\alpha}\right]^\alpha},$$

where A is a normalizing constant. This means that the optimizing Q^* is a mixture of the sources in the family, a result that is obtained in a more geometric and intuitive setting in [7]. Also note that if $\sum_{a\in\mathbb{X}} P_\theta(a)^{1/\alpha}$ is independent of θ, then $v = w$.

IV. Independent and Identically Distributed Sources

Consider now the problem of guessing the n-length realization of an independent and identically distributed (iid) source. Suppose that each component has a PMF P. Further suppose that guessing is done assuming a mismatched joint PMF. The results of the previous section are straightforwardly extended to such a source. The guessing exponent is normalized by n, and the additional terms go to 0 for large n, and are given by

$$\frac{1}{n}\log\left(1 + \ln|\mathbb{X}|^n\right) = O(\log n/n).$$

The results in the previous section on redundancy are for a finite number of sources in the family. However, Arikan and Merhav [3] show a *universal* guessing strategy that works for all iid sources with unknown P. Furthermore, the proof of their result indicates that the redundancy suffered by their universal guessing strategy is at most $O(\log n/n)$, for any iid source.

In the next section, we look at a more general class of independent, but not identically distributed set of sources.

V. Arbitrarily Varying Sources

Consider a finite state arbitrarily varying source (AVS) characterised by the PMF

$$P_{s^n}(x^n) = \prod_{i=1}^{n} P_{s_i}(x_i),$$

where x^n is the source sequence to be guessed, and $s^n = (s_1, \cdots, s_n)$ is an unknown arbitrary sequence of states. Each s_i takes values from a finite set $\mathbb{S}$.

This is a fairly large class of sources for which we would like to build a guessing strategy. We follow a hierarchical approach to keep the redundancy at a minimum.

Let U be a rational PMF on $\mathbb{S}$ where the rationals have denominator n. The set T_U is a subset of $\mathbb{S}^n$ whose elements have an empirical PMF equal to U. We then say that a sequence belonging to T_U has type T_U. The number of distinct types is upperbounded by $(n+1)^{|\mathbb{S}|}$ because the occurrence of each letter takes at most $n+1$ values.

Let us first assume that s^n belongs to T_U for some U. The following proposition, the main result of this paper, tells us how to arrive at a guessing strategy that gives the minmax penalty.

Proposition 3 *The normalized L_α-radius $K_n(T_U)$ for the family of sources with state sequence belonging to T_U is given by*

$$\begin{aligned} &K_n(T_U) \\ &\quad = \min_{Q} \max_{s^n\in T_U} \frac{1}{n} L_\alpha(P_{s^n}, Q) = \frac{1}{n}\left[H_\alpha(Q^*) - H_\alpha(P_{s^n})\right], \end{aligned}$$

where

$$Q^* = \sum_{s^n\in T_U} \frac{1}{|T_U|} P_{s^n}$$

is the uniform mixture.

Proof Outline: We first observe that for every $s^n \in T_U$, the quantity

$$\sum_{x^n\in\mathbb{X}^n} P_{s^n}(x^n)^\alpha$$

does not depend on the specific s^n. A simple permutation argument suffices to show this. This implies that the weights w that lead to the L_α-center are the same as the weights v that lead to the I_f-center, as observed in Section III.

Next, we check that the weights $w(s^n) = v(s^n) = 1/|T_U|$ satisfy the sufficient condition that $I_f(P'_{s^n}, Q'^*)$ is a constant independent of s^n (given that it belongs to T_U). To see this, we first write

$$\begin{aligned} I_f(P'_{s^n}, Q'^*) &= \sum_{x^n \in \mathbb{X}^n} P'_{s^n}(x^n)^{1+\rho} Q'^*(x^n)^{-\rho} \\ &= \frac{\sum_{x^n \in \mathbb{X}^n} P_{s^n}(x^n) Q'^*(x^n)^{-\rho}}{\left[\sum_{x^n \in \mathbb{X}^n} P_{s^n}(x^n)^{1/(1+\rho)}\right]^{1+\rho}} \quad (12) \end{aligned}$$

The quantity $Q'^*(x^n)$ is obtained from (10) and is given by

$$\begin{aligned} Q'^*(x^n) &= C_1 \left(\sum_{s^n \in T_U} \frac{1}{|T_U|} \left[P'_\theta(x^n) \right]^{1/\alpha} \right)^\alpha \\ &= C_2 \left(\sum_{s^n \in T_U} \frac{1}{|T_U|} P_\theta(x^n) \right)^\alpha . \end{aligned}$$

A permutation argument yields that this quantity depends on x^n only through the type of the sequence x^n. We can then use these facts to show that both the numerator and denominator of (12) do not depend on s^n, when $s^n \in T_U$. An evaluation then leads to the proposition. Details are available in [8]. □

The above Proposition along with Proposition 2, indicates that for every type T_U, and for any guessing strategy G, we can find a state sequence s^n for which

$$R(s^n, G) \geq K_n(T_U) - O(\log n/n).$$

We next consider the case when the type of the state sequence is not known. Within each type, we have found the penalty for a guessing strategy that depends only on the type of the state sequence. We then merge these lists according to the strategy used to prove Proposition 1. Since there are at most $(n+1)^{|\mathbb{S}|}$ types, this leads to an extra penalty of $O(\log n/n)$. We therefore have the following result.

Proposition 4 *For the finite-state AVS, there is a guessing strategy that does not depend on the state sequence, and whose penalty for a source with state sequence s^n belonging to T_U is upperbounded by $K_n(T_U) + O(\log n/n)$.*

Properties of $K_n(T_U)$ are under investigation.

REFERENCES

[1] J.L.Massey, “Guessing and entropy,” in *Proc. 1994 IEEE Int. Symp. on Information Theory* (Trondheim, Norway, 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] E.Arikan and N.Merhav, “Guessing subject to distortion ”, *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 1041-1056, vol. 44, no.3, May 1998.

[4] I.Csiszar, “A class of measures of informativity of observation channels,” *Periodica Mathematica Hungarica*, vol. 2, pp. 191-213, 1972.

[5] I-Csiszar, “Generalized cutoff rates and Rényi’s information measures”, *IEEE Trans. Inform. Theory*, vol. IT-41, pp. 26-34, Jan. 1995.

[6] R.Sibson, “Information radius”, *Z. Wahrscheinlichkeitstheorie verw. Geb.*, vol.14, pp. 149-160, 1969.

[7] R.Sundaresan,“A measure of discrimination and its geometric properties”, *Proceedings of the 2002 IEEE International Symposium on Information Theory*, p. 264, Lausanne, Switzerland, June 2002.

[8] R.Sundaresan, “Guessing under source uncertainty”, manuscript under preparation, December 2005.