Hypothesis testing via a comparator

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/ISIT.2012.6283845">http://dx.doi.org/10.1109/ISIT.2012.6283845</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers (IEEE)</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Fri Dec 21 16:13:13 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/78925">http://hdl.handle.net/1721.1/78925</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike 3.0</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/3.0/">http://creativecommons.org/licenses/by-nc-sa/3.0/</a></td>
</tr>
</tbody>
</table>
Hypothesis testing via a comparator

Yury Polyanskiy

Abstract—This paper investigates the best achievable performance by a hypothesis test satisfying a structural constraint: two functions are computed at two different terminals and the detector consists of a simple comparator verifying whether the functions agree. Such tests arise as part of study of fundamental limits of channel coding, but are also useful in other contexts. A simple expression for the Stein exponent is found and applied to showing a strong converse in the problem of multi-terminal hypothesis testing with rate constraints. Connections to the Gács-Körner common information and to spectral properties of conditional expectation operator are identified. Further tightening of results hinges on finding \( \lambda \) blocks of minimal weight. Application of Delsarte’s linear programming method to this problem is described.

I. INTRODUCTION

A classical problem in statistics and information theory is that of determining which of the two distributions, \( P \) or \( Q \), better fit an observed data vector. As shown by Neyman and Pearson, the binary hypothesis testing (in the case of simple hypotheses) admits an optimal solution based on thresholding the relative density of \( P \) with respect to \( Q \) (a Radon-Nikodym derivative). The asymptotic behavior of the tradeoff between the two types of errors has also been well studied by Stein, Chernoff, Hoeffding and Blahut. Knowledge of this tradeoff is important by itself and is also useful for other parts of information theory, such as channel coding [1, Section III.E] and data compression [2, Section IV.A].

The problem becomes, however, much more complex with the introduction of structural constraints on the allowable tests. For example, it may happen that observations consist of two parts, say \( X^n = (X_1, \ldots, X_n) \) and \( Y^n = (Y_1, \ldots, Y_n) \), which need to be compressed down to \( nR \) bits each before the decision is taken. Even the memoryless case, in which under either hypothesis the pairs \( (X_i, Y_i) \) are independent and identically distributed (i.i.d.) according to \( P_{XY} \) or \( Q_{XY} \), is a notoriously hard problem with only a handful of special cases solved [3]–[6]. Formally, this problem corresponds to finding the best test of the form

\[
T = 1\{ (f(X^n), g(Y^n)) \in A \},
\]

(1)

where optimization is over functions \( f \) and \( g \) with finite co-domains of cardinality \( 2^{nR} \) and critical regions \( A \). Here and below \( T = 1 \) designates the test choosing the distribution \( P \) and \( T = 0 \) the distribution \( Q \).

Another rich source of difficult problems is the distributed case, in which observations are taken by spatially separated sensors (whose measurements are typically assumed to be correlated in space but not in time). The goal is then to optimize the communication cost by designing (single letter) quantizers and a good (single or multi round) protocol for exchanges between the sensors and the fusion center; see [7]–[9] and references therein. These problems can again be restated in the form of constraining the allowable tests similar to (1).

In this paper we consider tests employing a comparator, namely those satisfying the constraint:

\[
T = 1\{ f(X^n) = g(Y^n) \},
\]

(2)

where the cardinality of the common co-domain of \( f \) and \( g \) is unrestricted. This constraint is motivated by the meta-converse method [1, Section III.E], which proves a lower bound on probability of error by first using a channel code as a binary hypothesis test and then comparing its performance with that of an optimal (Neyman-Pearson) test. However, so constructed test necessarily satisfies the structural constraint (2) and thus it is natural to investigate whether imposing (2) incurs exponential performance loss.

Another situation in which tests of the form (2) occur naturally is in the analysis of parallel systems, such as in fault-tolerant parallel computers, that under normal circumstances perform a redundant computation of a complicated function with high probability of agreement, while it is required to lower bound the probability of agreement when the fault occurs (modeled as \( P_{XY} \) changing to \( Q_{XY} \)). Yet another case is in testing hypotheses of biological nature based on the observation of zygosity of cells only (in eukaryotes).

The main result is that in the memoryless setting Stein exponent of tests satisfying (2) can indeed be quite a bit smaller than \( D(P_{XY} || Q_{XY}) \) and in fact is given by

\[
E = \min_{V_X = P_{X|V_Y} = P_Y} D(V_{XY} || Q_{XY}),
\]

(3)

where \( D(\|\cdot\|) \) is the Kullback-Leibler divergence, and the optimization is over all joint distributions \( V_{XY} \) with marginals matching those of \( P_{XY} \). In particular, \( E = 0 \) if (and only if) the marginals of \( Q_{XY} \) coincide with those of \( P_{XY} \).

In fact, for the latter case, the hypothesis testing with constraint (2) turns out to be intimately related to a problem of determining the common information \( C(X; Y) \) in the sense of Gács and Körner [10]. Using a technique pioneered by Witsenhausen [11] we show that the error probability cannot decay to zero at all (even subexponentially). Unfortunately, this is only shown under the condition that the confidence level is sufficiently high. Extending to the general case appears to be surprisingly hard. For a special case of binary \( X \) and \( Y \) we describe a bound based on Delsarte’s linear programming method [12] and demonstrate promising numerical results. However, we have not yet been able to identify a convenient method.

The author is with the Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139 USA, e-mail: yp@mit.edu.

The research was supported by the Center for Science of Information (CSO), an NSF Science and Technology Center, under grant agreement CCF-09-39370.
polynomial, such as found in [13] for the coding in Hamming space, admitting an asymptotic analysis.

The exponent $E$ has appeared before in the context of hypothesis testing with rate constraints (1), see [4, Theorems 5 and 8], and distributed detection [8, Theorem 2]. We identify the reasons for this below and also use this correspondence to prove the strong converse for the results in [4].

II. BACKGROUND AND NOTATION

Consider a distribution $P_{XY}$ on $X \times Y$. We denote a product distribution on $X^n \times Y^n$ by $P_{XY}^n$ and by $P_{XY} > 0$ the fact that $P_{XY}$ is non-zero everywhere on $X \times Y$.

Fix some $P_{XY}$ and $Q_{XY}$. For each integer $n \geq 1$ and $0 \leq \alpha \leq 1$ the performance of the best possible comparator hypothesis test of confidence level $\alpha$ is given by $\beta_{\alpha}(P_{XY}^n, Q_{XY}^n) \triangleq \inf \{ T \mid \mathbb{P}[T = 1] \geq \alpha \}$, where infimum is over all (perhaps, randomized) maps $f : X^n \rightarrow \mathbb{R}$ and $g : Y^n \rightarrow \mathbb{R}$ such that

$$\mathbb{P}[T = 1] \geq \alpha,$$

where $T$ is defined in (2). Here and below we follow the agreement that $\mathbb{P}$ and $\mathbb{Q}$ denote measures on some abstract spaces carrying random variables $(X^n, Y^n)$ distributed as $P_{XY}^n$ and $Q_{XY}^n$, resp.

For a finite $X \times Y$ and a given distribution $P_{XY}$ we define a bipartite graph with an edge joining $x \in X$ to $y \in Y$ if $P_{XY}(x, y) > 0$. The connected components of this graph are called components of $P_{XY}$ and the entropy of the random variable indexing the components is called the common information of $X$ and $Y$, cf. [10]. If the graph is connected, then $P_{XY}$ is called indecomposable. In particular indecomposability implies $P_X > 0$ and $P_Y > 0$.

We also define a maximal correlation coefficient $S(X; Y)$ between two random variables $X$ and $Y$ as

$$S(X; Y) = \sup_{f, g} \mathbb{E}[f(X)g(Y)],$$

supremum taken over all zero-mean functions of unit variance. For finite $X \times Y$ indecomposability of $P_{XY}$ implies $S(X; Y) < 1$ and (under assumption $P_X > 0, P_Y > 0$) is equivalent to it.

Finally, we recall [10] that a pair of sets $A \subseteq X^n$ and $B \subseteq Y^n$ is called a $\lambda$-block for $P_{XY}^n$ if $P_{XY}^n[A] > 0$, $P_{XY}^n[B] > 0$ and

$$\mathbb{P}[X^n \in A|Y^n \in B] \geq \lambda, \quad \mathbb{P}[Y^n \in B|X^n \in A] \geq \lambda.$$

An elegant theorem of Gács and Körner states

**Theorem 1 ([10]):** Let $P_{XY}$ be an indecomposable distribution on a finite $X \times Y$. Then for every $\lambda_n \geq \exp\{-o(n)\}$ there exists a sequence $\nu_n = o(n)$ such that for all $n$ any $\lambda_n$-block $(A, B)$ for $P_{XY}$ satisfies

$$P_{XY}^n[A \times B] \geq \exp\{-\nu_n\}.$$

III. MAIN RESULTS

A. Stein exponent

**Theorem 2:** Consider an indecomposable $P_{XY}$ on a finite $X \times Y$. Then for an arbitrary $Q_{XY}$ and any $0 < \alpha < 1$ we have

$$\lim_{n \to \infty} \frac{1}{n} \log \tilde{\beta}_{\alpha}(P_{XY}^n, Q_{XY}^n) = -E,$$

where $E$ is defined in (3). Moreover, if $E = \infty$ then there exists $n_0(\alpha)$ such that $\beta_{\alpha}(P_{XY}^n, Q_{XY}^n) = 0$ for all $n \geq n_0$.

**Proof:** Achievability: Consider functions

$$f(x^n) = 1\{x^n \notin T^n_{[P_X]}\},$$

$$g(y^n) = 2 \cdot 1\{y^n \notin T^n_{[P_Y]}\},$$

where $T^n_{[P]}$ denotes the set of $P$-typical sequences [14, Chapter 2] over the alphabet of $P$. Then, on one hand by typicality:

$$\mathbb{P}[f(x^n) = g(Y^n)] = P_{XY}^n[T^n_{[P_X]} \times T^n_{[P_Y]}] \geq 1 - o(1).$$

On the other hand, using joint-type decomposition it is straightforward to show that the set $T^n_{[P_X]} \times T^n_{[P_Y]}$ under the product measure $Q_{XY}^n$ satisfies

$$Q_{XY}^n[T^n_{[P_X]} \times T^n_{[P_Y]}] = \exp\{-nE + o(n)\}.$$  

Indeed, assuming otherwise we find a sequence of typical pairs $(x^n, y^n)$ with positive $Q_{XY}$-probability. But then the sequence of the joint types $V^n_{XY}$ associated to $(x^n, y^n)$ belongs to the closed set of joint distributions $\{V_{XY} : V_{XY} \ll Q_{XY}\}$ and by compactness must have a limit point $\hat{V}_{XY}$. By the $\delta$-convention [14, Chapter 2], the accumulation point must have marginals $\hat{V}_X = P_X$ and $\hat{V}_Y = P_Y$ and thus $E \leq D(\hat{V}_{XY}||Q_{XY}) < \infty$, a contradiction.

**Converse:** We reduce to the special case of the theorem, stated as Theorem 3 below. If $E = \infty$ then there is nothing to prove, so assume otherwise and take an arbitrary $V_{XY}$ with $V_X = P_X, V_Y = P_Y$ and $D(V_{XY}||Q_{XY}) < \infty$. Our goal is to show that

$$\tilde{\beta}_{\alpha}(P_{XY}^n, Q_{XY}^n) \geq \exp\{-nD(V_{XY}||Q_{XY}) + o(n)\}. \quad (9)$$

If $V_{XY} \neq 0$ then we can replace $V_{XY}$ with $(1 - \epsilon)V_{XY} + \epsilon P_X P_Y$, which is everywhere positive on $X \times Y$, and then take a limit as $\epsilon \to 0$ in (9). Thus we assume $V_{XY} > 0$.

Denote

$$A_n \triangleq \{ f(x^n) = g(Y^n) \}.$$  

By the special case of the theorem we have

$$V_{XY}^n[A_n] \geq \exp\{-o(n)\}. \quad (10)$$

Then, by a standard change of measure argument, we must have

$$Q_{XY}^n[A_n] \geq \exp\{-nD(V_{XY}||Q_{XY}) + o(n)\}. \quad (11)$$

Optimizing the choice of $V_{XY}$ in (11) proves (9) and the Theorem.

It remains to consider the case of matching marginals:

**Theorem 3 (Special case $E = 0$):** Let $P_{XY}$ be indecomposable, $Q_{XY} > 0$ and $Q_X = P_X, Q_Y = P_Y$. Then for any $0 < \alpha < 1$ we have

$$\tilde{\beta}_{\alpha}(P_{XY}^n, Q_{XY}^n) \geq \exp\{-o(n)\}. \quad (12)$$
Proof: First we show that any test of level \( \alpha \) must contain a \( \lambda \)-block with \( \lambda \geq \frac{\alpha}{2} \). Indeed, each pair \( \{ f(X^n) = i \}, \{ g(Y^n) = i \} \) is a \( \lambda \)-block for some \( \lambda_i \) (chosen to be maximum possible). Then, by the Bayes rule and \( \max\{x, y\} \leq x + y \) we get
\[
\mathbb{P}[f(X^n) = g(Y^n) = i] \leq \lambda_i(\mathbb{P}[f(X^n) = i] + \mathbb{P}[g(Y^n) = i]).
\]
Summing this over \( i \) shows that at least one \( \lambda_i \geq \frac{\alpha}{2} \).

By the Gács-Körner effect (Theorem 1) the probability of this \( \lambda \)-block is subexponentially large:
\[
\mathbb{P}[f(X^n) = g(Y^n) = i] \geq \exp\{-o(n)\}.
\]
Therefore, in particular we have (since the marginals of \( X^n \) and \( Y^n \) under \( \mathbb{P} \) and \( \mathbb{Q} \) coincide)
\[
\begin{align*}
Q[f(X^n) = i] & \geq \exp\{-o(n)\}, \quad (13) \\
Q[g(Y^n) = i] & \geq \exp\{-o(n)\}. \quad (14)
\end{align*}
\]
Thus, the sets \( \{ f(X^n) = i \} \) and \( \{ g(Y^n) = i \} \) must occupy a subexponential fraction of typical sets \( T^n_{P_X} \) and \( T^n_{P_Y} \). In view of (8) it is natural to expect that
\[
Q[f(X^n) = g(Y^n) = i] \geq \exp\{-o(n)\} \quad (15)
\]
(note that marginals match and thus \( E = 0 \) as per (3)). Under the assumption \( Q_{XY} > 0 \) it is indeed straightforward to show (15) by an application of blowing-up lemma; see [6, Theorem 3].

Finally, (15) completes the proof because
\[
Q[T = 1] \geq Q[f(X^n) = g(Y^n) = i].
\]

B. Discussion

It should be emphasized that although intuitively one imagines that the behavior of \( \tilde{\beta}_\alpha \) should markedly depend on how the connected components of \( P_{XY} \) and \( Q_{XY} \) relate to each other, Theorem 2 demonstrates that the Stein exponent is not sensitive to the decomposition of \( Q_{XY} \).

The assumption of indecomposability of \( P_{XY} \) in Theorem 2, however, is essential. Indeed, consider the case of \( X = Y = \{0, 1\} \) and \( X = Y \) uniform (under \( P_{XY} \)) vs \( X, Y \) independent uniform (under \( Q_{XY} \)). Clearly a test \( \{ X^n = Y^n \} \) demonstrates
\[
\tilde{\beta}_1(P^n_{XY}, Q^n_{XY}) \leq 2^{-n}, \quad (16)
\]
while according to the definition (3) we have \( E = 0 \).

We also remark that the case of \( E = \infty \) is possible. For example, let \( X, Y \) be binary with \( P_{XY}(0, y) = \frac{1}{2} - P_{XY}(1, y) = \frac{p}{2} \) for \( p > \frac{1}{2} \), and \( Q_{XY}(x, y) = \frac{1}{2} \{ x = y \} \).

C. Hypothesis testing with a 1-bit communication constraint

The exponent \( E \) in (3) is related to hypothesis testing under the communication constraint (1). In fact, Theorem 2 extends [4, Theorem 5] to the entire range \( 0 < \epsilon < 1 \), thereby establishing the full strong converse. This result has been obtained in [6] under different assumptions on \( P_{XY} \) and \( Q_{XY} \).

\[1\]Namely, we do not require \( D(P_{XY}||Q_{XY}) < \infty \) or positivity of \( Q_{XY} \), but require indecomposability of \( P_{XY} \).

Corollary 4: Consider a hypothesis testing between an indecomposable \( P_{XY} \) and an arbitrary \( Q_{XY} \) with structural restriction on tests of the form
\[
T = \{ (f(X^n), Y^n) \in A \} \quad (17)
\]
with binary-valued \( f \). Then for any \( 0 < \epsilon < 1 \) we have
\[
\inf Q[T = 1] = \exp\{-nE + o(n)\},
\]
where infimum is over all tests satisfying \( \mathbb{P}[T = 1] \geq 1 - \epsilon \) and \( E \) is given by (3).

Proof: Clearly, any test with binary-valued \( f \) and \( g \) of the form (2) is also a test of the form (17). Thus Theorem 2 establishes the achievability part. Conversely, for any test of the form (17) we may find sets \( A_0 \) and \( A_1 \) such that
\[
\mathbb{P}[T = 1] = \mathbb{P}\{Y^n \in A_0, f(X^n) = 0\} \\
\quad \cup \{Y^n \in A_1, f(X^n) = 1\} \geq 1 - \epsilon.
\]
Then without loss of generality assume that the first set in the union has \( \mathbb{P} \)-probability larger than \( \frac{1 - \epsilon}{2} \). Define the following function
\[
g(y^n) = 1\{y^n \in A_1 \setminus A_0\} + 21\{y^n \not\in A_0 \cup A_1\}
\]
Then since \( \{ f = g \} \supseteq \{ y^n \in A_0, f(X^n) = 0 \} \) we have
\[
\mathbb{P}[f(X^n) = g(Y^n)] \geq \frac{1 - \epsilon}{2},
\]
and thus by Theorem 2 we conclude that \( Q[T = 1] \) is at least \( \exp\{-nE + o(n)\} \).

We remark that the correspondence between the hypothesis tests with 1-bit compression and those of interest in this paper (2) does not hold in full generality. In particular, it was shown in [4, Theorem 5] that the exponent \( E \) in (3) is still optimal in the 1-bit scenario without the requirement of indecomposability of \( P_{XY} \), while example (16) demonstrates the contrary for our setup.

IV. NON-VANISHING LOWER BOUNDS

By Theorem 3 in the case when marginals of \( P_{XY} \) and \( Q_{XY} \) coincide the error cannot decay to zero exponentially. In fact, we conjecture that in the cases of matching marginals \( \tilde{\beta}_o(P^n_{XY}, Q^n_{XY}) \) does not vanish at all. In this section we prove the conjecture under additional assumptions and discuss potential methods for extending to the general case.
Theorem 5: Consider a \( P_{XY} \) and \( Q_{XY} = P_X P_Y \) such that \( S \triangleq S(X;Y) < \frac{1}{2} \) (under \( P_{XY} \)). Then for any \( \alpha \in (2S,1] \) we have
\[
\lim_{n \to \infty} \beta_\alpha (P^n_{XY}, Q^n_{XY}) > 0
\]

Proof: As in the proof of Theorem 3 given a test \( \{ f(X^n) = g(Y^n) \} \) of level \( \alpha \) we can extract a \( \lambda \)-block \((A,B)\) with \( \lambda > \frac{3}{4} \). We want to show that for some constant \( p = p(\alpha) > 0 \) and all \( n \) at least one of the marginals \( P^n_A [A] \) or \( P^n_B [B] \) can be bounded away from zero:
\[
\max \{ P^n_X [A], P^n_Y [B] \} \geq p
\]
Indeed, then we have
\[
Q(\mathcal{T} = 1) \geq P^n_X [A] P^n_Y [B] \geq \frac{\alpha}{2} p^2
\]
which follows because in a \( \lambda \)-block the smaller of the two probabilities in (21) should still be larger than the joint probability \( P^n_{XY} [A \times B] \) which is \( \geq \lambda p \). Finally, the estimate (21) follows from the next result.

Lemma 6: Consider a \( \lambda \)-block \((A,B)\) for \( P^n_{XY} \). Then,
\[
\max \{ P^n_X [A], P^n_Y [B] \} \geq \min \left\{ 1, \frac{\lambda - S}{2}, \frac{1}{1-S} \right\}
\]
whenever \( S = S(X;Y) < 1 \).

Proof: Consider an operator \( T_n : L_2(Y^n, P^n_Y) \to L_2(X^n, P^n_X) \) defined as follows:
\[
(T_n h)(x^n) \triangleq \mathbb{E} [f(Y^n)|X^n = x^n],
\]
where the expectation is over the distribution \( P^n_{XY} \). Note that the second largest singular value of \( T_n \) is precisely the maximal correlation coefficient \( S(X;Y) \) (under \( P_{XY} \)), see [15]. Thus, for any zero-mean functions \( h \in L_2(X^n) \) and \( h' \in L_2(Y^n) \) we have
\[
\mathbb{E} [h(X^n)h'(Y^n)] = (T_n h', h) \leq S(X;Y) \|h\|_2 \|h'\|_2.
\]
Denote \( p_A = P^n_X[A] \), \( p_B = P^n_Y[B] \) and assume \( p_B \geq p_A \). If \( p_B \geq \frac{1}{2} \) then there is nothing to prove, so assume otherwise. Then we have
\[
\lambda p_B \leq P^n_X [A \times B] \leq p_A p_B + S \sqrt{p_A(1-p_A)p_B(1-p_B)} \leq p_B^2 + S p_B(1-p_B),
\]
where \( (26) \) is by the definition of a \( \lambda \)-block, \( (27) \) is by \( (25) \) applied to \( h(x^n) = 1 \{ x^n \in A \} - p_A h \) and \( h'(y^n) = 1 \{ y^n \in B \} - p_B h' \); and \( (28) \) is because \( p_A \leq p_B \leq \frac{1}{2} \). Canceling \( p_B \) on both sides in \( (28) \) we obtain \( (23) \).

Next, we discuss what is required to extend Theorem 5 to full generality. To handle a general \( Q_{XY} \) one needs a non-vanishing lower bound independent of \( n \) on
\[
\lambda_{\min} (p, Q^n_{XY}) = \min_{A,B} \lambda^n_{XY} [A \times B],
\]
where the minimization is over \( Q^n_X[A], Q^n_Y[B] \geq p \). For \( Q_{XY} = P_X P_Y \) this problem is void since \( \lambda_{\min} (p, P^n_{XY}) = p^2 \). Nevertheless, even the case of \( Q_{XY} = P_X P_Y \) is far from being resolved as we need to extend to the full range \( 0 < \alpha < 1 \). We discuss this second problem further.

A. More on spectral methods

In a nutshell, the proof of Theorem 5 consisted of two steps. First, we identified a Markov chain
\[
F \to X^n \to Y^n \to G,
\]
where we denoted \( F \triangleq f(X^n) \), \( G \triangleq g(Y^n) \). Note that by the data-processing for maximal correlation we have
\[
S(F;G) \leq S(X^n;Y^n) = S(X;Y).
\]
Second, for large \( \alpha \) we showed a lower bound
\[
Q[F = G] \geq \sum_{i} P[F = i] P[G = i] \geq \text{const} > 0
\]
under conditions: a) \( \mathbb{P}[F = G] \geq \alpha \) and b) \( S(F;G) \leq S \). Can a lower bound be tightened so that it does not vanish for all \( \alpha > 0 \)?

The answer is negative. Indeed, consider a distribution \( P_{FG} \) on \([M] \times [M] \):
\[
P_{FG}(i,j) = \frac{\alpha}{M} 1\{i = j\} + \frac{1 - \alpha}{M(M-1)} 1\{i \neq j\}
\]
Then we have \( S(F;G) = \alpha - \frac{1 - \alpha}{M} \). That is, such \( P_{FG} \) satisfies the \( \alpha \)-constraint and the maximal correlation constraint whenever \( \alpha \leq S(X;Y) \) and achieves
\[
\sum_i P[F = i] P[G = i] = \frac{1}{M} \to 0
\]
as \( M \to \infty \).

It may appear that as a workaround one may consider higher spectral invariants in addition to \( S(X;Y) \). Formally, to any joint distribution \( P^n_{XY} \) we associate the operator \( T_n \) in (24). Let the singular values of \( T_n \) sorted in decreasing order be
\[
1 = \sigma_{n,0} \geq \sigma_{n,1} \geq \sigma_{n,2} \geq \cdots \geq 0,
\]
where \( \sigma_{n,1} = S(X;Y) \). Since \( T_n = T^\otimes n \) the singular spectrum of \( T_n \) consists of all possible products of the form \( \prod_{t=1}^n \sigma_{1,j_t} \), and in particular
\[
1 = \sigma_{n,0} \geq \sigma_{n,1} = \cdots = \sigma_{n,n} = S(X;Y).
\]
Moreover, it is easy to show that if one has a Markov chain (29) then singular values \( \{ \mu_j, j = 1, \ldots \} \) associated with \( P_{FG} \) are related to those of \( P^n_{XY} \) via the following "spectral-processing" inequalities:
\[
\prod_{j=1}^k \mu_j \leq \prod_{j=1}^k \sigma_{n,j} \quad k = 1, \ldots
\]
Clearly this extends the data-processing for maximal correlation used in the proof of Theorem 5. Does it lead to a lower-bound non-vanishing for all \( \alpha \)?

Alas, the answer is negative. Indeed, in the example (30) the singular spectrum associated to \( P_{FG} \) consists of 1 and \( \frac{1}{M-1} \) (of multiplicity \( M-1 \)). This spectrum satisfies (31) as long as \( \alpha \leq S(X;Y) \) and \( M \leq n+1 \). Thus, for \( \alpha \leq S(X;Y) \), inequalities (31) can not rule out the possibility that
\[
Q[F = G] \leq \frac{1}{n+1}.
\]
B. \( \lambda \)-blocks of minimal weight

Another method to extend the range of \( \alpha \) in Theorem 5 is to find a non-vanishing (as \( n \to \infty \)) lower bound on the marginal probability \( P_X \circ [A] \) of a \( \lambda \)-block \( (A, B) \). In fact, it is enough to consider the case of \( P_{XY} \) with \( X = Y \) and \( P_X = P_Y \). Indeed, consider an arbitrary \( \lambda \)-block \( (A, B) \) and construct a Markov kernel \( W : \mathcal{X} \to \mathcal{X} \) as composition \( W = P_{X|Y} \circ P_{Y|X} \), namely

\[
W(x_1|x_0) = \sum_{y \in \mathcal{Y}} P_{X|Y}(x_1|y)P_{Y|X}(y|x_0).
\]

Then distribution \( P_X \) is a stationary distribution of the Markov chain associated with \( W \) (and operator of conditional expectation (24) is self-adjoint). Moreover, we clearly have

\[
W^n(A|A) \geq P^n_X(A|A),
\]

where maximum is over all non-negative \( \{\alpha_k\} \) such that \( \alpha_0 = 1 \), \( \beta_0(\alpha) \geq 0 \) and (34).

To give the dual formulation of (35) say that a polynomial \( P(x) \) of degree not larger than \( n \) is admissible if

\[
P(x) = \sum_{v=0}^n p_v K_v(x),
\]

and \( p_v \geq (1 - 2\delta)^v \) for all \( v = 0, \ldots, n \). Then, we have

\[
\lambda_n^*(p) \leq \min \left( 2^{-n} P(0) + (p - 2^{-n}) \max_{x=1,\ldots,n} P(x) \right),
\]

where minimum is over all admissible polynomials. The bound of Lemma 6 states

\[
\lambda_n^*(p) \leq 1 + 2\delta(p - 1),
\]

and corresponds to choosing

\[
P(x) = K_0(x) + (1 - 2\delta) \sum_{v=1}^n K_v(x)
\]

As numerical evaluation of (35) shows, see Fig. 1, the bound (36) can be significantly improved. Finding a suitable admissible polynomial \( P(x) \) remains an open problem.

**References**


