Non-Typewise Method for Sharper Upper Bounds

Gautam Dasarathy

In this note, we outline an idea that one can use to tighten bounds obtained using the method of types. This is based on problems 7,11 and 12 from Chapter 2 of Csiszar and Korner's book.

A brief note about notation before we proceed. $\mathcal{X}$ denotes a finite set and let $\mathbf{x} \in \mathcal{X}^k$ denotes the length k vector $(x_1, x_2, \ldots, x_k)$ from $\mathcal{X}^k$. Given a bounded function M on $\mathcal{X}$, $M^k(\mathbf{x}) := \prod_{i=1}^k M(x_i)$ and for any $F \subset \mathcal{X}^k$, $M^k(F) := \sum_{\mathbf{x} \in F} M^k(\mathbf{x})$. $\mathbb{1}\{\cdot\}$ denotes the indicator function. Further, we use standard notation for information theoretic quantities.

The main result can be stated as follows.

Theorem 1. *Let us suppose that we have a bounded "mass" function $M(\cdot)$ on $\mathcal{X}$. Then, for any $F \in \mathcal{X}^k$, the following holds*

$$M^k(F) \leq \exp\left\{-kD(P_{M,F} \| M)\right\} \tag{1}$$

where $P_{M,F}(a) := \sum_{\mathbf{x} \in F} \frac{M^k(\mathbf{x})}{M^k(F)} P_{\mathbf{x}}(a)$, for all $a \in \mathcal{X}$ and $D(P_{M,F} \| M) := \mathbb{E}_{P_{M,F}}\left[\log\left(\frac{P_{M,F}}{M}\right)\right]$.

Proof. To prove this, we define a random vector $X^k = (X_1, X_2, \ldots, X_k)$ which is drawn according to the following distribution

$$\Pr\left\{X^k = \mathbf{x}\right\} := \frac{M^k(\mathbf{x})}{M^k(F)} \mathbb{1}\left\{\mathbf{x} \in F\right\} \tag{2}$$

and independently of $J \overset{\text{unf}}{\sim} \{1, 2, \ldots, k\}$. First, we upper bound the joint entropy of X^k by the sum of individual entropies and proceed as follows

$$\begin{aligned}
H(X_1, X_2, \ldots, X_k) &\leq \sum_{i=1}^k H(X_i) \\
&= k \sum_{i=1}^k \frac{1}{k} H(X_i) \\
&= kH(X_J | J) \\
&\leq kH(X_J)
\end{aligned}$$

Now observe that $P(X_J = a) = \sum_{i=1}^k \frac{1}{k} \sum_{\mathbf{x}} P(X^k = \mathbf{x}) = P_{M,F}(a)$. This gives us the following upper bound

$$H(X_1, \ldots, X_k) \leq -k \sum_{a \in \mathcal{X}} P_{M,F}(a) \log P_{M,F}(a) \tag{3}$$

Alternatively, the above joint entropy can be evaluated as

$$H(X_1, X_2, \ldots, X_k) = -\sum_{\mathbf{x}\in F} \frac{M^k(\mathbf{x})}{M^k(F)} \log\left(\frac{M^k(\mathbf{x})}{M^k(F)}\right) \tag{4}$$

$$= \log M^k(F) - \sum_{\mathbf{x}\in F} \frac{M^k(\mathbf{x})}{M^k(F)} \log M^k(\mathbf{x}) \tag{5}$$

$$= \log M^k(F) - \sum_{\mathbf{x}\in F} \frac{M^k(\mathbf{x})}{M^k(F)} k \sum_{a\in\mathcal{X}} P_{\mathbf{x}}(a) \log M(a) \tag{6}$$

$$= \log M^k(F) - k \sum_{a\in\mathcal{X}} P_{M,F}(a) \log M(a) \tag{7}$$

(3) and (7) together conclude the proof. □

Corollary 1. *For any $F \in \mathcal{X}^k$, the following bound on the size of F holds*

$$|F| \le \exp\{kH(P_{1,F})\} \tag{8}$$

where $P_{1,F}(a) := \frac{1}{|F|}\sum_{\mathbf{x}\in F} P_{\mathbf{x}}(a)$. Further, for any distribution Q on $\mathcal{X}$, we have

$$Q^k(F) \le \exp\{-kD(P_{Q,F}\|Q)\} \tag{9}$$

where $P_{Q,F}(a) := \sum_{\mathbf{x}\in F} \frac{Q^k(\mathbf{x})}{Q^k(F)} P_{\mathbf{x}}(a)$ for all $a \in \mathcal{X}$.

Proof. For the first part, set $M(a) = 1$ for all $a \in \mathcal{X}$ and the second part follows by setting $M(a) = Q(a)$ for all $a \in \mathcal{X}$. □

This corollary tells us that the size of an arbitrary set in $\mathcal{X}^k$ can be bounded in terms of the entropy of the "average type of the sequences of that set" and that a similar statement holds for $Q^k(F)$.

To see why these results are useful, we now consider three examples. In each of these cases, standard type-based arguments would give us exponential bounds which are tight only as $k \to \infty$. But, using the "non-typewise" bounding of Theorem 1, we show that under some conditions, the same bounds hold non-asymptotically.

1. **Error Exponent for Binary Block Codes**. We now show that for any finite set $\mathcal{X}$ and rate $R > 0$, there exists a k−to−n_k block code such that for any DMS with alphabet $\mathcal{X}$ and arbitrary distribution P, the probability of error satisfies

$$P_e \le \exp\left\{-k \min_{Q:H(Q)\ge R} D(Q\|P)\right\} \tag{10}$$

Observe that this bound does not involve a polynomial factor as is usual in proofs by the method of types. To see this , let $A_k := \bigcup_{Q:H(Q)<R} \mathcal{T}_Q$. The encoding function essentially maps one-to-one from A_k to an integer from $\{1, 2, \ldots, 2^{kR}\}$ and anything in $\mathcal{X}^k \setminus A^k$ is mapped to 1 (say). By defining $Q(\cdot) = \sum_{\mathbf{x}\in A_k^c} \frac{P^k(\mathbf{x})}{P(A_k^c)} P_{\mathbf{x}}(\cdot)$, we can use the results of Corollary 1 to get

$$P\left(\mathcal{X}^k \setminus A^k\right) = P\left(\left\{\mathbf{x} \in \mathcal{X}^k : H(P_{\mathbf{x}}) \ge R\right\}\right) \tag{11}$$

$$\le \exp\{-kD(Q\|P)\}. \tag{12}$$

(10) follows directly from this since, by the concavity of entropy, we know that $H(Q) \ge R$.

2. **Sanov's Theorem**. Let $\mathscr{P}$ be a set of distributions on the alphabet $\mathcal{X}$ and let Q be another distribution on $\mathcal{X}$. Sanov's theorem gives us a bound on the probability that a random sample drawn according to Q would "appear as though it was drawn from a distribution in $\mathscr{P}$". This bound is $(k+1)^{|\mathcal{X}|} \exp\{-k \inf_{P \in \mathscr{P}} D(P\|Q)\}$. However, if $\mathscr{P}$ is a *convex* set of distributions, then the following tighter bound holds

$$\frac{1}{k} \log Q^k \left(\left\{ \mathbf{x} \in \mathcal{X}^k : P_{\mathbf{x}} \in \mathscr{P} \right\} \right) \leq - \inf_{P \in \mathscr{P}} D\left(P \| Q\right) \tag{13}$$

and this also follows as a direct consequence of Corollary 1.

3. **Hypothesis Testing**. Following along the same lines, we can get a stronger result for the probability of missed detection in hypothesis testing. We can actually show the following statement:

For any given P and $a > 0$, there exists $A_k \subset \mathcal{X}^k$ such that

$$\lim_{k \to \infty} \frac{1}{k}(1 - P^k(A_k)) = -a \tag{14}$$

and for every Q

$$\frac{1}{k} \log Q^k(A_k) \leq - \min_{\hat{P}: D(\hat{P}\|P) \leq a} D(\hat{P}\|P) \tag{15}$$