From the book
Imre Csiszár – János Körner: Information Theory*
Coding Theorems for Discrete Memoryless Systems

2 Types and typical sequences

Most of the proof techniques used in this book will be based on a few simple combinatorial lemmas, summarized below.

Drawing k times independently with distribution Q from a finite set X, the probability of obtaining the sequence $\mathbf{x} \in \mathsf{X}^k$ depends only on how often the various elements of X occur in $\mathbf{x}$. In fact, denoting by $N(a|\mathbf{x})$ the number of occurrences of $a \in \mathsf{X}$ in $\mathbf{x}$, we have

$$Q^k(\mathbf{x}) = \prod_{a \in \mathsf{X}} Q(a)^{N(a|\mathbf{x})}. \tag{2.1}$$

DEFINITION 2.1 The *type* of a sequence $\mathbf{x} \in \mathsf{X}^k$ is the distribution $P_{\mathbf{x}}$ on X defined by

$$P_{\mathbf{x}}(a) \triangleq \frac{1}{k} N(a|\mathbf{x}) \quad \text{for every} \quad a \in \mathsf{X}.$$

For any distribution P on X, the set of sequences of type P in X^k is denoted by T_P^k or simply T_P. A distribution P on X is called a *type of sequences in* X^k if $\mathsf{T}_P^k \neq \emptyset$. ○

Sometimes the term "type" will also be used for the sets $\mathsf{T}_P^k \neq \emptyset$ when this does not lead to ambiguity. These sets are also called *type classes* or *composition classes*.

REMARK In mathematical statistics, if $\mathbf{x} \in \mathsf{X}^k$ is a sample of size k consisting of the results of k observations, the type of $\mathbf{x}$ is called the *empirical distribution* of the sample $\mathbf{x}$. ○

By (2.1), the Q^k-probability of a subset of T_P is determined by its cardinality. Hence the Q^k-probability of any subset A of X^k can be calculated by combinatorial counting arguments, looking at the intersections of A with the various sets T_P separately. In doing so, it will be relevant that the number of different types in X^k is much smaller than the number of sequences $\mathbf{x} \in \mathsf{X}^k$.

LEMMA 2.2 (*Type counting*) The number of different types of sequences in X^k is less than $(k+1)^{|\mathsf{X}|}$. ○

→ 2.1

Proof For every $a \in \mathsf{X}$, $N(a|\mathbf{x})$ can take $k+1$ different values. □

The next lemma explains the role of entropy from a combinatorial point of view, via the asymptotics of a multinomial coefficient.

*Second Edition, Cambridge University Press, 2011.

LEMMA 2.3 For any type P of sequences in X^k → 2.2

$$(k+1)^{-|\mathsf{X}|}\exp[kH(P)] \leqq |\mathsf{T}_P| \leqq \exp[kH(P)].$$ ○

Proof Since (2.1) implies

$$P^k(\mathbf{x}) = \exp[-kH(P)] \quad \text{if} \quad \mathbf{x} \in \mathsf{T}_P$$

we have

$$|\mathsf{T}_P| = P^k(\mathsf{T}_P)\exp[kH(P)].$$

Hence it is enough to prove that

$$P^k(\mathsf{T}_P) \geqq (k+1)^{-|\mathsf{X}|}.$$

This will follow by the Type counting lemma if we show that the P^k-probability of $\mathsf{T}_{\widehat{P}}$ is maximized for $\widehat{P} = P$.

By (2.1) we have

$$P^k(\mathsf{T}_{\widehat{P}}) = |\mathsf{T}_{\widehat{P}}| \cdot \prod_{a\in\mathsf{X}} P(a)^{k\widehat{P}(a)} = \frac{k!}{\prod_{a\in\mathsf{X}}(k\widehat{P}(a))!}\prod_{a\in\mathsf{X}} P(a)^{k\widehat{P}(a)}$$

for every type $\widehat{P}$ of sequences in X^k.

It follows that

$$\frac{P^k(\mathsf{T}_{\widehat{P}})}{P^k(\mathsf{T}_P)} = \prod_{a\in\mathsf{X}} \frac{(kP(a))!}{(k\widehat{P}(a))!} P(a)^{k(\widehat{P}(a)-P(a))}.$$

Applying the obvious inequality $\frac{n!}{m!} \leqq n^{n-m}$, this gives

$$\frac{P^k(\mathsf{T}_{\widehat{P}})}{P^k(\mathsf{T}_P)} \leqq \prod_{a\in\mathsf{X}} k^{k(P(a)-\widehat{P}(a))} = 1.$$ □

If X and Y are two finite sets, the *joint type* of a pair of sequences $\mathbf{x} \in \mathsf{X}^k$ and $\mathbf{y} \in \mathsf{Y}^k$ is defined as the type of the sequence $\{(x_i, y_i)\}_{i=1}^k \in (\mathsf{X}\times\mathsf{Y})^k$. In other words, it is the distribution $P_{\mathbf{x},\mathbf{y}}$ on $\mathsf{X}\times\mathsf{Y}$ defined by

$$P_{\mathbf{x},\mathbf{y}}(a,b) \triangleq \frac{1}{k}N(a,b|\mathbf{x},\mathbf{y}) \quad \text{for every} \quad a\in\mathsf{X}, b\in\mathsf{Y}.$$

Joint types will often be given in terms of the type of $\mathbf{x}$ and a stochastic matrix $V : \mathsf{X} \to \mathsf{Y}$ such that

$$P_{\mathbf{x},\mathbf{y}}(a,b) = P_{\mathbf{x}}(a)V(b|a) \quad \text{for every} \quad a\in\mathsf{X}, b\in\mathsf{Y}. \tag{2.2}$$

Note that the joint type $P_{\mathbf{x},\mathbf{y}}$ uniquely determines $V(b|a)$ for those $a \in \mathsf{X}$ which do occur in the sequence $\mathbf{x}$. For conditional probabilities of sequences $\mathbf{y} \in \mathsf{Y}^k$, given a sequence $\mathbf{x} \in \mathsf{Y}^k$, the matrix V of (2.2) will play the same role as the type of $\mathbf{y}$ does for unconditional probabilities.