Data Processing Lemma. Let $M$, $X$, and $Y$ be as in Example 8.2. Then

$$I(M \wedge \varphi(Y)) \leq I(X \wedge \varphi(Y)) \leq I(X \wedge Y).$$

Proof. If we can show that $M \to X \to \varphi(Y)$ is a Markov chain, then $I(M \wedge \varphi(Y)) \leq I(X \wedge \varphi(Y))$. Since we have already seen in Example 8.1 that $X \to Y \to \varphi(Y)$ is a Markov chain, we have $I(X \wedge \varphi(Y)) \leq I(X \wedge Y)$.

The fact that $M \to X \to \varphi(Y)$ is a Markov chain follows immediately from the previous example by putting $B_j := \{y : \varphi(y) = j\}$ and writing

$$P(\varphi(Y) = j | X = x, M = i) = P(Y \in B_j | X = x, M = i) = W_n(B_j | x).$$

Proof of Fano’s Inequality. Consider the random variable $Z := I_{\{M \neq V\}}$. Using the law of total probability,

$$H(M|V) := E \left[ \log \frac{1}{P(M|V)} \right] = E \left[ E \left[ \log \frac{1}{P(M|V)} | Z \right] \right] = E \left[ \log \frac{1}{P(M|V)} | Z = 1 \right] P(M \neq V) + E \left[ \log \frac{1}{P(M|V)} | Z = 0 \right] P(M = V).$$

Next, by Jensen’s inequality, the conditional expectations satisfy

$$E \left[ \log \frac{1}{P(M|V)} | Z = 1 \right] \leq \log E \left[ \frac{1}{P(M|V)} | Z = 1 \right]$$

and

$$E \left[ \log \frac{1}{P(M|V)} | Z = 0 \right] \leq \log E \left[ \frac{1}{P(M|V)} | Z = 0 \right].$$

It remains to evaluate the two conditional expectations on the right. Observe that

$$P(M = m, V = v | M \neq V) = \frac{P(M = m, V = v)}{P(M \neq V)} I_{\{m\}^c}(v) = \frac{P(m|v)P(v)}{P(M \neq V)} I_{\{m\}^c}(v).$$

$^b$ We write $P(m|v)$ for $P_{M|V}(m|v)$ and $P(v)$ for $P_V(v)$ to simplify the notation.
and 
\[ P(M = m, V = v | M = V) = \frac{P(M = m, V = v)}{P(M = V)} I_{\{m\}}(v) = \frac{P(m|V)P(v)}{P(M = V)} I_{\{m\}}(v). \]

Hence,
\[
E\left[ \frac{1}{P(M|V)} \right] | Z = 1 = \sum_v \sum_m \frac{1}{P(m|V)} \frac{P(m|V)P(v)}{P(M \neq V)} I_{\{m\}}(v) \\
= \frac{1}{P(M \neq V)} \sum_v P(v) \sum_m I_{\{m\}}(v) = \frac{N - 1}{P(M \neq V)}.
\]

Similarly,
\[
E\left[ \frac{1}{P(M|V)} \right] | Z = 0 = \sum_v \sum_m \frac{1}{P(m|V)} \frac{P(m|V)P(v)}{P(M = V)} I_{\{m\}}(v) \\
= \frac{1}{P(M = V)} \sum_v P(v) \sum_m I_{\{m\}}(v) = \frac{1}{P(M = V)}.
\]

Putting this all together yields
\[
H(M|V) \leq P(M \neq V) \log \frac{N - 1}{P(M \neq V)} + P(M = V) \log \frac{1}{P(M = V)} \\
= P(M \neq V) \log(N - 1) + h(P(M \neq V)).
\]

\[\square\]

### 8.3. The Weak Converse

If \( R \) is an achievable rate for a DMC \( W \) under the average probability of error criterion, then \( R \leq \sup_p I(p \times W) \).

**Proof.** Fix any \( 0 < \lambda < 1 \) and any \( \Delta R > 0 \). Since \( R \) is achievable, for all sufficiently large \( n \), there exist codewords \( x_1, \ldots, x_N \) and there exists a decoder \( \varphi \) with \( \frac{\log N}{n} > R - \Delta R \) and \( P(\varphi(Y) \neq M) < \lambda \), where \( M, X, Y \) are as in Example 8.2 with \( W_n = W^n \) and \( q(i) = \frac{1}{N} \); i.e.,
\[
P(M = i, X = x, Y = y) = \frac{1}{N} I_{\{x_i\}}(x) W^n(y|x).
\]

We can then write
\[
\log N = H(M) \leq \lambda \log N + \log 2 + I(M \land \varphi(Y)),
\]
from which it follows by the corollary to Fano’s inequality that

\[(1 - \lambda) \log N \leq I(M \land \varphi(Y)) + \log 2.\]

Since \((\log N)/n > R - \Delta R\), we have \(\log N > n(R - \Delta R)\), and so

\[(1 - \lambda)n(R - \Delta R) \leq I(M \land \varphi(Y)) + \log 2 \leq I(X \land Y) + \log 2,\]

by the Data Processing Lemma. We now observe that

\[
I(X \land Y) = H(Y) - H(Y|X) \\
= H(Y) - \sum_{k=1}^{n} H(Y_k|X_k), \quad \text{by the DMC assumption,} \\
\leq \sum_{k=1}^{n} H(Y_k) - \sum_{k=1}^{n} H(Y_k|X_k) \\
= \sum_{k=1}^{n} I(X_k \land Y_k) \\
= \sum_{k=1}^{n} I(p_k \times W) \\
\leq n \cdot \sup_{p} I(p \times W),
\]

Hence,

\[(1 - \lambda)n(R - \Delta R) \leq n \cdot \sup_{p} I(p \times W) + \log 2,\]
or

\[R \leq \frac{\sup_{p} I(p \times W)}{1 - \lambda} + \frac{\log 2}{n(1 - \lambda)} + \Delta R,
\]

and it follows that \(R \leq \sup_{p} I(p \times W).\)