Proof of Th. 3.5:

(a) g.m.d. code with codeword lengths $n_1, \ldots, n_m$.
Let $r = \max \{ n_i \}$ max. codeword length.

$\beta = \{ \beta_{ij} | n_i = l \}$ no. of codewords of length $l \in \mathbb{N}$, $l \leq r$

It holds, $k \in \mathbb{N}$

$$\left( \sum_{j=1}^{m} d^{-n_j} \right)^k = \left( \sum_{l=1}^{r} \beta_{l} d^{-l} \right)^k = \sum_{l=k}^{k \cdot r} y_{l} \cdot d^{-l}$$

with

$$y_{l} = \sum_{1 \leq i_1 < \ldots < i_k \leq r} \beta_{i_1} \ldots \beta_{i_k} \quad l = k \cdot \ldots \cdot k \cdot r$$

$y_{l}$: no of source words of length $k$ which have codeword length $l$.

$d^{l}$: no of all codewords of length $l$.

Since $g$ is m.d., each codeword has at most one sourceword. Hence

$$y_{l} \leq d^{l}$$

$$\left( \sum_{j=1}^{m} d^{-n_j} \right)^k \leq \sum_{l=k}^{k \cdot r} d^{l} d^{-l} = k \cdot r - k + 1 \leq k \cdot r \quad \forall k \in \mathbb{N}.$$ 

Further

$$\sum_{j=1}^{m} d^{-n_j} \leq (k \cdot r)^{\frac{1}{k} \to 1} \quad (k \to \infty)$$

so that

$$\sum_{j=1}^{m} d^{-n_j} \leq 1.$$
Huffman are optimal, i.e., have shortest average codeword length. We consider the case $d=2$.

**Lemma A.**

Let $X = \{x_1, \ldots, x_m\}$ with probabilities $p_1 \geq \ldots \geq p_m > 0$. There exists an optimal binary prefix code $g$ with codeword lengths $\eta_1, \ldots, \eta_m$ such that

1. $\eta_1 \leq \ldots \leq \eta_m$
2. $\eta_{m-1} = \eta_m$
3. $g(x_{m-1})$ and $g(x_m)$ differ only in the last position.

**Proof.** Let $g$ be an optimum prefix code with $\eta_1, \ldots, \eta_m$.

(i) If $p_i > p_j$ then necessarily $\eta_i \leq \eta_j$, $1 \leq i < j \leq m$. Otherwise exchange $g(x_i)$ and $g(x_j)$ to obtain a code $g'$ with

$$\bar{w}(g') - \bar{w}(g) = p_i \eta_j + p_j \eta_i - p_i \eta_i - p_j \eta_j$$

$$= (p_i - p_j)(\eta_j - \eta_i) < 0$$

contradicting optimality of $g$.

(ii) There is an opt. prefix code $\tilde{g}$ with $\eta_1 \leq \ldots \leq \eta_m$. If $\eta_{m-1} < \eta_m$ delete $\eta_m - \eta_{m-1}$ positions of $g(x_m)$ to obtain a better code.

(iii) If $l_1 \leq \ldots \leq l_{m-1} = l_m$ for an opt. prefix code $g$, and $g(x_{m-1})$ and $g(x_m)$ differ by more than the last position, delete the last position in both to obtain a better code. \( \square \)
Lemma B.
Let \( X = \{x_1, \ldots, x_m\} \) with prob. \( p_1 \geq \ldots \geq p_m > 0 \).
\( X' = \{x'_1, \ldots, x'_{m-1}\} \) with prob. \( p'_i = p_i, \ i = 1, \ldots, m-2 \)
\[ p'_{m-1} = p_{m-1} + p_m \]

\( g' \) an opt. prefix code for \( X' \) with

codewords \( g'(x'_i), i = 1, \ldots, m-1 \).

Then
\[ g(x_i) = \begin{cases} g'(x'_i), & i = 1, \ldots, m-2 \\ g'(x'_{m-1}), & i = m-1 \\ g'(x'_{m-1}), & i = m \end{cases} \]
is an optimal prefix code for \( X \).

Proof. Denote codeword lengths \( n_i, n'_i \) for \( g, g' \) respectively.

\[ \bar{n}(g) = \sum_{j=1}^{m-1} p_j n_j + (p_m + p_{m-1}) (n_{m-1} + 1) \]
\[ = \sum_{j=1}^{m-2} p_j n'_j + p_{m-1} (n'_{m-1} + 1) \]
\[ = \sum_{j=1}^{m-1} p_j n'_j + p_{m-1} + p_m = \bar{n}(g') + p_{m-1} + p_m \]

Assume \( g \) is not optimal for \( X \). There exists an opt. prefix code \( c' \) with properties (i)-(iii) of Lemma and
\[ \bar{n}(c') < \bar{n}(g) \]
Set
\[ h'(x_j) = \begin{cases} h(x_j), & j = 1, \ldots, m-2 \\ L h(x_{m-1}) \end{cases} \]
by deleting the last position of \( h(x_{m-1}), j = m \)

Then \( \bar{n}(h') + p_{m-1} + p_m = \bar{n}(h) < \bar{n}(g) = \bar{n}(g') + p_{m-1} + p_m \)

Hence \( \bar{n}(h') < \bar{n}(g') \) contradicting optimality of \( g' \).