Proof of Th. 3.5:

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

$\beta_\ell = |\{ i \mid n_i = \ell \}|$ no. of codewords of length $\ell \in \mathbb{N}$, $\ell \le r$

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

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

with

$$\gamma_\ell = \sum_{\substack{1 \le i_1, \ldots, i_k \le r \\ i_1 + \ldots + i_k = \ell}} \beta_{i_1} \cdots \beta_{i_k}, \quad \ell = k, \ldots, k \cdot r$$

γ_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 u.d., each codeword has at most one sourceword. Hence

$$\gamma_\ell \le d^\ell.$$

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

Further

$$\sum_{j=1}^{m} d^{-n_j} \le (kr)^{1/k} \longrightarrow 1 \quad (k \to \infty)$$

so that $\sum_{j=1}^{m} d^{-n_j} \le 1$. $\blacksquare$

Huffman are optimal, i.e., have shortest average codeword length. We consider the case $d=2$.

Lemma A.

Let $\mathcal{X} = \{x_1, \dots, x_m\}$ with probabilities $p_1 \geq \dots \geq p_m > 0$.
There exists an optimal binary prefix code g with codeword lengths $n_1, \dots, n_m$ such that

(i) $n_1 \leq \dots \leq n_m$

(ii) $n_{m-1} = n_m$

(iii) $g(x_{m-1})$ and $g(x_m)$ differ only in the last position. ⌟

Proof. Let g be an optimum prefix code with $n_1, \dots, n_m$.

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

$$\bar{n}(g') - \bar{n}(g) = p_i n_j + p_j n_i - p_i n_i - p_j n_j$$
$$= (p_i - p_j)(n_j - n_i) < 0$$

contradicting optimality of g.

(ii) There is an opt. prefix code g with $n_1 \leq \dots \leq n_m$.
If $n_{m-1} < n_m$ delete $n_m - n_{m-1}$ positions of $g(x_m)$ to obtain a better code.

(iii) If $l_1 \leq \dots \leq l_{m-1} = l_m$ 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. ∎

Lemma B.

Let $\mathcal{X} = \{x_1, \dots, x_m\}$ with prob. $p_1 \geq \dots \geq p_m > 0$.

$\mathcal{X}' = \{x_1', \dots, x_{m-1}'\}$ with prob. $p_i' = p_i$, $i = 1, \dots, m-2$

$p_{m-1}' = p_{m-1} + p_m$

g' an opt. prefix code for $\mathcal{X}'$ with codewords $g'(x_i')$, $i = 1, \dots, m-1$.

Then

$$g(x_i) = \begin{cases} g'(x_i'), & i = 1, \dots, m-2 \\ (g'(x_{m-1}'), 0), & i = m-1 \\ (g'(x_{m-1}'), 1), & i = m \end{cases}$$

is an optimal prefix code for $\mathcal{X}$.

Proof. Denote codeword lengths n_i, n_i' for g, g', respectively.

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

$$= \sum_{j=1}^{m-2} p_j' n_j' + p_{m-1}' (n_{m-1}' + 1)$$

$$= \sum_{j=1}^{m-1} p_j' n_j' + p_{m-1} + p_m = \bar{n}(g') + p_{m-1} + p_m$$

Assume g is not optimal for $\mathcal{X}$. There exists an opt. prefix code h with properties (i)–(iii) of La A. and $\bar{n}(h) < \bar{n}(g)$.

Set

$$h'(x_j') = \begin{cases} h(x_j), & j = 1,\dots,m-2 \\ \lfloor h(x_{m-1}) \rfloor \text{ by deleting the last position of } h(x_{m-1}), & j = m \end{cases}$$

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'. $\boxtimes$