Solution to Exercise 16.

Theorem 4.3 shall be proven.

(a) $X$ is a discrete random variable with $p_i = P(X = x_i), i = 1, \ldots, m$. It holds

$$H(X) = - \sum_i p_i \log p_i \geq 0,$$

as $p_i \geq 0$ and $-\log p_i \geq 0$ for $0 < p_i \leq 1$ and $0 \cdot \log 0 = 0$ per definition. Equality holds, if all addends are zero, i.e.,

$$p_i \log p_i = 0 \iff p_i \in \{0, 1\}, i = 1, \ldots, m,$$

as $p_i > 0$ and $-\log p_i > 0$, thus, $-p_i \log p_i > 0$ for $0 < p_i < 1$.

(b)

$$H(X) - \log m = - \sum_i p_i \log p_i - \sum_{i = 1}^m p_i \log m$$

$$= \sum_{i : p_i > 0} p_i \log \frac{1}{m p_i}$$

$$= (\log e) \sum_{i : p_i > 0} p_i \ln \frac{1}{m p_i}$$

$$\leq (\log e) \sum_{i : p_i > 0} p_i \left( \frac{1}{m p_i} - 1 \right)$$

$$= (\log e) \left( \sum_{i : p_i > 0} \frac{1}{m} - 1 \right) \leq 0.$$

As $\ln z = z - 1$ only holds for $z = 1$ it follows that equality holds iff $p_i = 1/m, i = 1, \ldots, m$. In particular, it follows $p_i > 0, i = 1, \ldots, m$.

(c) Define for $i = 1, \ldots, m$ and $j = 1, \ldots, d$

$$p_{ij} = P(X = x_i \mid Y = y_j).$$
Show \( H(X \mid Y) - H(X) \leq 0 \) which is equivalent to the claim.

\[
H(X \mid Y) - H(X) = -\sum_{i,j} p_{i,j} \log p_{i,j} + \sum_i p_i \log p_i
\]

\[
= -\sum_{i,j} p_{i,j} \log \frac{p_{i,j}}{p_j} + \sum_i \sum_j p_{i,j} \log p_i
\]

\[
= \log(e) \sum_{i,j: p_{i,j} > 0} p_{i,j} \ln \frac{p_i p_j}{p_{i,j}}
\]

\[
\leq \log(e) \sum_{i,j: p_{i,j} > 0} p_{i,j} \left( \frac{p_i p_j}{p_{i,j}} - 1 \right)
\]

\[
= \log(e) \left( \sum_{i,j: p_{i,j} > 0} p_i p_j - 1 \right) \leq 0
\]

Note that from \( p_{i,j} > 0 \) it follows \( p_i, p_j > 0 \). Equality hold for \( \frac{p_i p_j}{p_{i,j}} = 1 \) which is equivalent to \( X \) and \( Y \) being stochastically independent.

This means that the transinformaton \( I(X,Y) = H(X) - H(X \mid Y) \) is nonnegative.

(d) It holds

\[
H(X, Y) = -\sum_{i,j} p_{i,j} \log p_{i,j}
\]

\[
= -\sum_{i,j} p_{i,j} \left[ \log p_{i,j} - \log p_i + \log p_i \right]
\]

\[
= -\sum_{i,j} p_{i,j} \log \frac{p_{i,j}}{p_i} - \sum_i \sum_j p_{i,j} \log p_i
\]

\[
= H(Y \mid X) + H(X).
\]

(e) It holds

\[
H(X,Y) \overset{(d)}{=} H(X) + H(Y \mid X) \overset{(c)}{=} H(X) + H(Y)
\]

with equality as in (c) iff \( X \) and \( Y \) are stochastically independent.