Problem 2.1:

(a) X follows a geometric distribution: $P\{X = k\} = (1-p)^{k-1}\, p$, $k = 1, 2, 3, \ldots$, where $p = 1/2$. Therefore

$$\begin{aligned} H(X) &= -\sum_{k=1}^{\infty} p\,(1-p)^{k-1} \log_2 p\,(1-p)^{k-1} \\ &= -\sum_{k=1}^{\infty} \frac{1}{2^k} \log_2 \frac{1}{2^k} \\ &= \sum_{k=1}^{\infty} k \left(\frac{1}{2}\right)^k \\ &= \frac{1/2}{(1-1/2)^2} = 2 \text{ bits.} \end{aligned}$$

(b) One possible "efficient" series of questions is: "Is $X = 1$?"; "If not, is $X = 2$?"; "If not, is $X = 3$?"; ...; etc. The resulting expected number of questions is equal to

$$\sum_{n=1}^{\infty} n \left(\frac{1}{2}\right)^n = 2 = H(X).$$

This reinforces the intuition that $H(X)$ is a measure of the uncertainty of X. In general, $E[\#\text{ of questions}] \geq H(X)$.

Problem 2.2: Let X be discrete R.V. with finite alphabet $\mathcal{X}$ and let $Y = f(X)$. In general, we have $H(X) \geq H(f(X))$ with equality holding *iff* $f(\cdot)$ is one-to-one. To show this,

$$\begin{aligned} H(X, f(X)) &= H(X) + H(f(X)|X) \\ &= H(X) \\ &= H(f(X)) + \underbrace{H(X|f(X))}_{\geq 0}. \end{aligned}$$

Hence, $H(X) \geq H(f(X))$ with equality *iff* X is a function of Y; i.e., $f(\cdot)$ is one-to-one.
(a) For $Y = 2^X$, $H(X) = H(Y)$ since 2^X is one-to-one.
(b) For $Y = \cos X$, $H(Y) \leq H(X)$ since $\cos(\cdot)$ is *not* one-to-one, in general.

Problem 2.3: $H(X) = \sum_{x \in \mathcal{X}} -p(x) \log_2 p(x) \geq 0$. Each term in this summation is non-negative. Thus, $H(X) = 0$ iff each term is zero, which can only happen if $p(x) = 0 or 1$. Hence, $H(X)$ is zero iff X is deterministic.

Problem 2.5:

(a) $H(X, g(X)) = H(X) + H(g(X)|X)$ by the chain rule for entropies.

(b) $H(g(X)|X) = 0$ since for any given value of X, $g(X)$ is fixed since $g(\cdot)$ is a deterministic function of X.

(c) $H(X, g(X)) = H(g(X)) + H(X|g(X))$, again by the chain rule.

(d) Since $H(X|g(X)) \geq 0$ with equality iff X is a function of $g(X)$ (i.e., $g(\cdot)$ is invertible).

Problem 2.6: Assume $H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X=x) = 0$. Let $\hat{\mathcal{X}} = \{x \in \mathcal{X} : p(x) > 0\}$. Then

$$H(Y|X) = \sum_{x \in \hat{\mathcal{X}}} p(x) H(Y|X=x) = 0.$$

Each term in the sum above is non-negative and $p(x) > 0$. So $H(Y|X=x) = 0 \forall x \in \hat{\mathcal{X}}$. This means that $\forall x \in \hat{\mathcal{X}}$, there is some y for which $p(y|x) = 1$. Thus Y is a (deterministic) function of X.

Problem 2.8: The two teams play until one of them wins 4 games. Assuming both teams are equally matched, the distribution of X is:

- There are 2 world series (AAAA, BBBB) with 4 games. Each occurs with probability $(\frac{1}{2})^4 = 1/16$.
- There are $2\binom{4}{3} = 8$ world series with 5 games; each one occurring with probability $(\frac{1}{2})^5 = 1/32$.
- There are $2\binom{5}{3} = 20$ world series with 6 games; each one occurring with probability $(\frac{1}{2})^6 = 1/64$.
- There are $2\binom{6}{3} = 40$ world series with 7 games; each one occurring with probability $(\frac{1}{2})^7 = 1/128$.

So

$$\begin{aligned} H(X) &= \sum_x p(x) \log_2 \frac{1}{p(x)} \\ &= 2(\frac{1}{16}) \log_2 16 + 8(\frac{1}{32}) \log_2 32 + 20(\frac{1}{64}) \log_2 64 + 40(\frac{1}{128}) \log_2 128 \\ &= 5.813 \text{ bits.} \end{aligned}$$

The distribution of Y is:

- $Pr(Y=4) = 2(\frac{1}{16}) = 1/8$.
- $Pr(Y=5) = 8(\frac{1}{32}) = 1/4$.
- $Pr(Y=6) = 20(\frac{1}{64}) = 5/16$.
- $Pr(Y=7) = 40(\frac{1}{128}) = 5/16$.

So

$$H(Y) = \sum_{y=4}^{7} p(y) \log_2 \frac{1}{p(y)} = 1.924 \quad bits$$

$H(Y|X) = 0$ since Y is a deterministic function of X (if we know X, we know Y). Now,

$$\begin{aligned} H(X,Y) &= H(X) + H(Y|X) = H(X) \\ &= H(Y) + H(X|Y), \end{aligned}$$

by the chain rule. Hence,

$$\begin{aligned} H(X|Y) &= H(X) - H(Y) \\ &= 5.813 - 1.924 = 3.889 \text{ bits.} \end{aligned}$$

Problem 2.10:

(a) Take (X,Y) of Problem 2.16. Here, $I(X;Y) = 0.252$. Let $Z = X$. $I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = 0$. So this is an example in which $I(X;Y) > I(X;Y|Z)$.

(b) Take $Z = X + Y$ where X and Y are i.i.d. with $p_X(0) = p_X(1) = p_Y(0) = p_Y(1) = 1/2$. Then $I(X;Y) = 0$ by independence. Now

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z).$$

The second term ($H(X|Y,Z)$) must be zero since X is a function of Y and Z ($X = Z - Y$).

Now for the first term. Z can be 0, 1 or 2. But there is uncertainty in X only when $Z = 1$ (we know $X = 0$ when $Z = 0$ and we know $X = 1$ when $Z = 2$). Hence,

$$I(X;Y|Z) = H(X|Z) = \sum_{z=0}^{2} p_Z(z)H(X|Z=z) = 0 + (1/2)\log_2 2 + 0 = 1/2\text{bits}.$$

So this is an example in which $I(X;Y) < I(X;Y|Z)$.

Problem 2.16:

(a) $P\{X = 0\} = 2/3$; hence $H(X) = h_b(2/3) = (2/3)\log_2 3/2 + (1/3)\log_2 3 = 0.918$ bits. $P\{Y = 0\} = 1/3$; hence $H(Y) = h_b(1/3) = H(X) = 0.918$ bits.

(b) $H(X|Y) = (1/3)H(X|Y=0) + (2/3)H(X|Y=1) = (1/3)*0 + (2/3)*1 = 2/3$ bits.
$H(Y|X) = (2/3)H(Y|X=0) + (1/3)H(Y|X=1) = (2/3)*1 + (1/3)*0 = 2/3$ bits.

(c) $H(X,Y) = H(X) + H(Y|X) = 1.5847$ bits.

(d) $I(X:Y) = H(Y) - H(Y|X) = 0.2513$ bits.

(e) Same result as in (d).

(f) See lecture notes.

Problem 2.18: Let X be a discrete RV with alphabet $\mathcal{X} = \{x_1, \ldots, x_r\}$ and Y be a discrete RV with alphabet $\mathcal{Y} = \{y_1; \ldots, y_s\}$. Define $Z = X + Y$.

(a) $P(Z = z|X = x) = P(X + Y = z|X = x) = P(Y = z - x|X = x)$. Hence

$$\begin{aligned}
H(Z|X) &= \sum_x p(x)H(Z|X=x) \\
&= -\sum_x p(x)\sum_z p(Z=z|X=x)\log_2 p(Z=z|X=x) \\
&= -\sum_x p(x)\sum_z p(Y=z-x|X=x)\log_2 p(Y=z-x|X=x) \\
&= \sum_x p(x)H(Y|X=x) \\
&= H(Y|X).
\end{aligned}$$

Since X and Y are independent, $H(Y|X) = H(Y)$. Hence, $H(Y) = H(Z|X)$.

Now, $H(Z) \geq H(Z|X) = H(Y)$, where the inequality follows from the fact that conditioning reduces entropy. Thus, $H(Z) \geq H(Y)$.

Similarly, $H(Z) \geq H(Z|Y) = H(X)$. Hence $H(Z) \geq H(X)$.

(b) Let X be a binary uniform RV and let $Y = -X$, then $H(X) = H(Y) = 1$ and $Z = 0$ with probability one. So $H(Z) = 0$ and $H(X) = H(Y) > H(Z)$.

(c) Since $Z = X + Y$, Z is a function of (X, Y) and thus $H(Z) \leq H(X,Y)$ (by the results from Problem 2.5). But $H(X,Y) =\leq H(X) + H(Y)$ by the independence bound. So

$$H(Z) \overset{(1)}{\leq} H(X,Y) \overset{(2)}{\leq} H(X) + H(Y)$$

where equality holds in (1) iff (X, Y) is a function of Z; i.e., $Z = X + Y$ is invertible; and equality holds in (2) iff X and Y are independent.

Problem 2.28: Define

$$\begin{aligned} P_1 &\triangleq (p_1, \ldots, p_i, \ldots, p_j, \ldots, p_m), \\ P_2 &\triangleq (p_1, \ldots, \frac{p_i + p_j}{2}, \ldots, \frac{p_i + p_j}{2}, \ldots, p_m). \end{aligned}$$

Now

$$\begin{aligned} H(P_2) - H(P_1) &= -2\left(\frac{p_i + p_j}{2}\right)\log_2\left(\frac{p_i + p_j}{2}\right) + p_i \log_2 p_i + p_j \log_2 p_j \\ &= -(p_i + p_j)\log_2\left(\frac{p_i + p_j}{2}\right) + p_i \log_2 p_i + p_j \log_2 p_j \\ &\geq 0, \end{aligned}$$

where the last step follows directly from the log-sum inequality:

$$\sum_{k=1}^{2} a_k \log_2 (a_k/b_k) \geq (\sum_{k=1}^{2} a_k) \log_2 ((\sum_{k=1}^{2} a_k)/(\sum_{k=1}^{2} b_k))$$

with $a_1 = p_i$, $a_2 = p_j$ and $b_1 = b_2 = 1$.
In general, it can similarly be shown (using the log-sum inequality) that if, for any $1 \leq s \leq m$,

$$\begin{aligned} P_1 &\triangleq (p_1, \ldots, p_{i_1}, \ldots, p_{i_s}, \ldots, p_m), \\ P_2 &\triangleq (p_1, \ldots, \frac{p_{i_1} + \cdots + p_{i_s}}{s}, \ldots, \frac{p_{i_1} + \cdots + p_{i_s}}{s}, \ldots, p_m), \end{aligned}$$

then

$$H(P_2) \geq H(P_1).$$

Problem B:

$$\begin{aligned} H(p_1, p_2, \ldots, p_n) &= -\sum_{i=1}^{n} p_i \log p_i \\ &= -\sum_{i}^{m} p_i \log p_i - \sum_{i=m+1}^{n} p_i \log p_i. \end{aligned}$$

But by the log-sum inequality,

$$\begin{aligned} \sum_{i=m+1}^{n} p_i \log p_i &\geq \underbrace{\left(\sum_{i=m+1}^{n} p_i\right)}_{q} \log \frac{(\sum_{i=m+1}^{n} p_i)}{(\sum_{i=m+1}^{n} 1)} \\ &= q \log q - q \log(n-m). \end{aligned}$$

So,

$$\begin{aligned} H(p_1, \ldots, p_n) &\leq -\sum_{i=1}^{m} p_i \log p_i - q \log q + q \log(n-m) \\ &= H(p_1, p_2, \ldots, p_m, q) + q \log(n-m) \end{aligned}$$

With equality *iff* $p_i = \frac{q}{n-m}$, $\forall\, i = m+1, \ldots, n$. □

Please let me know if you see a mistake in this solution.