Sample problems for the 4th lecture (Mar. 22)

1. Let the joint distribution of X and Y be:

$P_{X,Y}(\cdot,\cdot)$	$y=0$	$y=1$	$y=2$
$x=0$	$\frac{1}{8}$	0	$\frac{1}{4}$
$x=1$	$\frac{1}{8}$	$\frac{1}{8}$	0
$x=2$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

(a) Compute

$$H(X), H(Y), H(X|Y), H(Y|X), H(X,Y) \text{ and } I(X;Y),$$

and indicate the quantities (in bits) for each area of the Venn diagram.

(b) Let $Z = \begin{cases} 0, & y=2; \\ 1, & y \in \{0,1\} \end{cases}$. Find $I(X;Z)$ and compare it with $I(X;Y)$ to verify the data processing inequality.

(c) Let $\hat{X} := g(Y)$ be the optimal MAP estimator of X from observing Y. Define the probability of error as

$$P_e := \Pr[\hat{X} \neq X].$$

Check whether equality holds for Fano's inequality:

$$H(X|Y) \leq h_{\mathrm{b}}(P_e) + P_e \cdot \log_2(|\mathcal{X}| - 1),$$

where $h_{\mathrm{b}}(x) := -x\log_2 x - (1-x)\log_2(1-x)$ for $0 \leq x \leq 1$ is the binary entropy function.

(d) Can we find a lousy estimator such that the resulting error probability is 1? Justify your answer.

Hint: Check whether $H(X|Y)$ lies between $\log_2(|\mathcal{X}| - 1)$ and $\log_2(|\mathcal{X}|)$.

Solution.

(a)

$$\begin{aligned} H(X;Y) &= \mathbb{E}\left[\log_2 \frac{1}{P_{X,Y}(X,Y)}\right] \\ &= \sum_{x=0}^{2}\sum_{y=0}^{2} P_{X,Y}(x,y)\cdot \log_2 \frac{1}{P_{X,Y}(x,y)} \\ &= 6\cdot\frac{1}{8}\log_2\frac{1}{1/8}+\frac{1}{4}\log_2\frac{1}{1/4} \\ &= \frac{11}{4} \end{aligned}$$

$$\begin{aligned} H(X) &= \mathbb{E}\left[\log_2 \frac{1}{P_X(X)}\right] \\ &= 2\cdot\frac{3}{8}\log_2\frac{1}{3/8}+\frac{1}{4}\log_2\frac{1}{1/4} \\ &= \frac{11}{4}-\frac{3}{4}\log_2(3) \end{aligned}$$

$$\begin{aligned} H(Y) &= \mathbb{E}\left[\log_2 \frac{1}{P_Y(Y)}\right] \\ &= 2\cdot\frac{3}{8}\log_2\frac{1}{3/8}+\frac{1}{4}\log_2\frac{1}{1/4} \\ &= \frac{11}{4}-\frac{3}{4}\log_2(3) \end{aligned}$$

With the knowledge of $H(X,Y)$, $H(X)$ and $H(Y)$, we can apply the chain rule for entropy to obtain:

$$H(X|Y) = H(X,Y) - H(Y) = \frac{3}{4}\log_2(3)$$

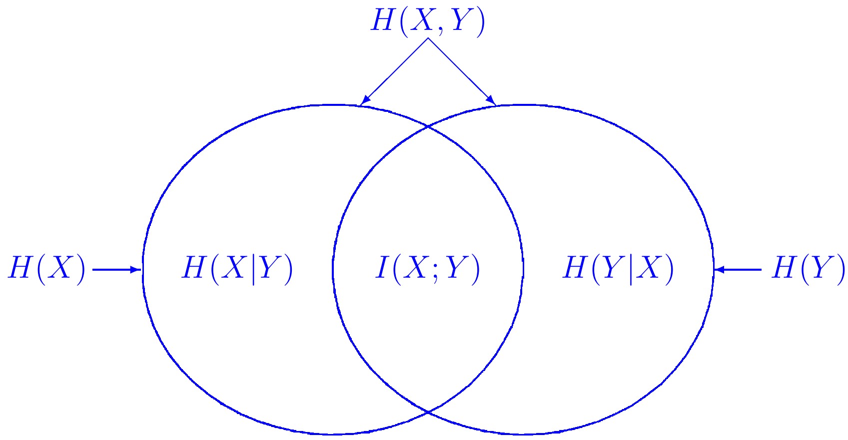
and

$$H(Y|X) = H(X,Y) - H(X) = \frac{3}{4}\log_2(3)$$

Last, by definition of mutual information, we obtain:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) = \frac{11}{4} - \frac{3}{2}\log_2(3)$$

You shall indicate the quantities (in bits) for each area of the Venn diagram by yourself.



(b) With

$P_{X,Z}(\cdot,\cdot)$	$z=0$	$z=1$
$x=0$	$\frac{1}{4}$	$\frac{1}{8}$
$x=1$	0	$\frac{1}{4}$
$x=2$	$\frac{1}{8}$	$\frac{1}{4}$

we derive

$$\begin{aligned} H(X,Z) &= 2\cdot\frac{1}{8}\log_2\frac{1}{1/8} + 3\cdot\frac{1}{4}\log_2\frac{1}{1/4} \\ &= \frac{9}{4} \end{aligned}$$

and

$$\begin{aligned} H(Z) &= \frac{3}{8}\log_2\frac{1}{3/8} + \frac{5}{8}\log_2\frac{1}{5/8} \\ &= 3 - \frac{3}{8}\log_2(3) - \frac{5}{8}\log_2(5), \end{aligned}$$

which implies

$$I(X;Z) = H(X) + H(Z) - H(X,Z) = 2 - \frac{3}{8}\log_2(3) + \frac{5}{8}\log_2(5).$$

Comparing $I(X;Y)$ and $I(X;Z)$, we examine

$$I(X;Y) - I(X;Z) = \frac{7}{2} - \frac{9}{8}\log_2(3) - \frac{5}{8}\log_2(5) \approx 0.266 \text{ bits} > 0,$$

which confirms the validity of the data processing inequality.

(c) The MAP estimator is given by

$$g(y) = \begin{cases} 0, & y = 2; \\ \text{arbitrary in } \{1,2\}, & y = 1; \\ \text{arbitrary in } \{0,1,2\}, & y = 0 \end{cases}$$

In fact, we can set $y(y) = z$ in (b). Hence, $P_e = \Pr[X \neq Z] = \frac{1}{2}$. As a result, Fano's inequality becomes

$$\frac{3}{4}\log_2(3) \leq h_b\left(\frac{1}{2}\right) + \frac{1}{2}\log_2(3-1) = \frac{3}{2}$$
$$\Leftrightarrow \quad \log_2(3) \leq 2 \Leftrightarrow 3 \leq 2^2 = 4.$$

Hence, equality in Fano's inequality cannot be achieved in this case.

(d) When $H(X|Y) = \frac{3}{4}\log_2(3)$ lies between $\log_2(|\mathcal{X}|-1) = 1$ and $\log_2(|\mathcal{X}|) = \log_2(3)$, Fano's inequality also provides an upper bound to the estimation error. Hence, no estimator can result in the lousy error performance of 1.

Note: In fact, the estimator that has the worst performance is the *minimum a posteriori* estimator, i.e.,

$$g_{\text{worst}}(y) = \begin{cases} 1, & y = 2; \\ 0, & y = 1; \\ \text{arbitrary in } \{0,1,2\}, & y = 0 \end{cases}$$

We can choose

$$g_{\text{worst}}(y) = \begin{cases} 1, & y = 2; \\ 0, & y \in \{0,1\} \end{cases}$$

With

$P_{X,g_{\text{worse}}(Y)}(\cdot,\cdot)$	$g_{\text{worst}}(y) = 0$	$g_{\text{worst}}(y) = 1$
$x = 0$	$\frac{1}{8}$	$\frac{1}{4}$
$x = 1$	$\frac{1}{4}$	0
$x = 2$	$\frac{1}{4}$	$\frac{1}{8}$

we obtain that the worst performance is $7/8$, and Fano's inequality becomes

$$\frac{3}{4}\log_2(3) \leq h_b\left(\frac{7}{8}\right) + \frac{7}{8}\log_2(3-1) = \frac{31}{8} - \frac{7}{8}\log_2(7)$$
$$\Leftrightarrow \quad 2\log_2(3) + 7\log_2(7) \approx 22.82 \leq 31.$$

Again, Fano's inequality provides a strict "overbound" in this case.

2. (a) Consider random variables X and Y with alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, where $\mathcal{X}$ is finite and $\mathcal{Y}$ can be countably infinite, and assume that for each $x \in \mathcal{X}$, we are given a *ternary partition* $\{\mathcal{S}_x, \mathcal{T}_x, \mathcal{V}_x\}$ on the observation space $\mathcal{Y}$. Define

$$p := \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{S}_x} P_{X,Y}(x,y), \; q := \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{T}_x} P_{X,Y}(x,y), \; r := \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{V}_x} P_{X,Y}(x,y)$$

and

$$s := \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{S}_x} P_Y(y), \quad t := \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{T}_x} P_Y(y), \quad v := \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{V}_x} P_Y(y).$$

Show that

$$H(X|Y) \le H(p,q,r) + p\log_2(s) + q\log_2(t) + r\log_2(v), \qquad (1)$$

where

$$H(p,q,r) = p\log_2\frac{1}{p} + q\log_2\frac{1}{q} + r\log_2\frac{1}{r}.$$

Give the necessary and sufficient condition under which equality in (1) holds.

Hint: Subtract one side from the other side and apply the fundamental inequality.

(b) Show that Fano's inequality is a special case of (1) by specifying $\{\mathcal{S}_x, \mathcal{T}_x, \mathcal{V}_x\}$.

Solution.

(a)

$$
\begin{aligned}
&H(X|Y) - H(p,q,r) - p\log_2(s) - q\log_2(t) - r\log_2(v) \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{S}_x} P_{X,Y}(x,y)\log_2\frac{1}{P_{X|Y}(x|y)} + \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{T}_x} P_{X,Y}(x,y)\log_2\frac{1}{P_{X|Y}(x|y)} \\
&\quad + \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{V}_x} P_{X,Y}(x,y)\log_2\frac{1}{P_{X|Y}(x|y)} + \left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{S}_x} P_{X,Y}(x,y)\right]\log_2\left(\frac{p}{s}\right) \\
&\quad + \left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{T}_x} P_{X,Y}(x,y)\right]\log_2\left(\frac{q}{t}\right) + \left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{V}_x} P_{X,Y}(x,y)\right]\log_2\left(\frac{r}{v}\right) \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{S}_x} P_{X,Y}(x,y)\log_2\frac{p}{P_{X|Y}(x|y)\cdot s} + \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{T}_x} P_{X,Y}(x,y)\log_2\frac{q}{P_{X|Y}(x|y)\cdot t} \\
&\quad + \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{V}_x} P_{X,Y}(x,y)\log_2\frac{r}{P_{X|Y}(x|y)\cdot v} \\
&\le \log_2(e)\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{S}_x} P_{X,Y}(x,y)\left[\frac{p}{P_{X|Y}(x|y)\cdot s} - 1\right] \\
&\quad + \log_2(e)\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{T}_x} P_{X,Y}(x,y)\left[\frac{q}{P_{X|Y}(x|y)\cdot t} - 1\right] \\
&\quad + \log_2(e)\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{V}_x} P_{X,Y}(x,y)\left[\frac{r}{P_{X|Y}(x|y)\cdot v} - 1\right] \\
&= \log_2(e)\left[\frac{p}{s}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{S}_x} P_Y(y) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{S}_x} P_{X,Y}(x,y)\right] \\
&\quad + \log_2(e)\left[\frac{q}{t}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{T}_x} P_Y(y) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{T}_x} P_{X,Y}(x,y)\right] \\
&\quad + \log_2(e)\left[\frac{r}{v}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{V}_x} P_Y(y) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{V}_x} P_{X,Y}(x,y)\right] \\
&= \log_2(e)\left[\frac{p}{s}(s) - p\right] + \log_2(e)\left[\frac{q}{t}(t) - q\right] + \log_2(e)\left[\frac{r}{v}(v) - r\right] \\
&= 0,
\end{aligned}
$$

where the inequality follows from the FI Lemma. Equality holds iff

$$P_{X|Y}(x|y) = \begin{cases} \dfrac{p}{s}, & y \in \mathcal{S}_x \\ \dfrac{q}{t}, & y \in \mathcal{T}_x \\ \dfrac{r}{v}, & y \in \mathcal{V}_x \end{cases}$$

(b) Let $\mathcal{S}_x = \{y \in \mathcal{Y} : g(y) = x\}$, $\mathcal{T}_x = \mathcal{S}_x^c$ and $\mathcal{V}_x = \emptyset$, where $g(\cdot)$ is an estimator. Then, $p = 1 - P_e$, $q = P_e$, $r = 0$,

$$\begin{aligned} s &:= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{S}_x} P_Y(y) \\ &= \sum_{x\in\mathcal{X}} \sum_{y:g(y)=x} P_Y(y) \\ &= \sum_{y\in\mathcal{Y}} \sum_{x:g(y)=x} P_Y(y) \\ &= \sum_{y\in\mathcal{Y}} P_Y(y) \cdot |\{x \in \mathcal{X} : g(y) = x\}| \\ &= \sum_{y\in\mathcal{Y}} P_Y(y) \\ &= 1 \end{aligned}$$

and $t = |\mathcal{X}| - s = |\mathcal{X}| - 1$. Accordingly,

$$\begin{aligned} H(X|Y) &\leq H(p,q,r) + p\log_2(s) + q\log_2(t) + r\log_2(v) \\ &= h_{\mathrm{b}}(P_e) + (1-P_e)\log_2(1) + P_e\log_2(|\mathcal{X}|-1) + 0\log_2(0) \\ &= h_{\mathrm{b}}(P_e) + P_e\log_2(|\mathcal{X}|-1) \end{aligned}$$

with equality holding iff

$$P_{X|Y}(x|y) = \begin{cases} 1 - P_e, & g(y) = x \\ \dfrac{P_e}{|\mathcal{X}|-1}, & g(y) \neq x \end{cases}$$

3. Define the *divergence typical set* as

$$\mathcal{A}_n(\delta) := \left\{ x^n \in \mathcal{X}^n : \left| \frac{1}{n} \log_2 \frac{P_{X^n}(x^n)}{P_{\hat{X}^n}(x^n)} - D(P_X \| P_{\hat{X}}) \right| < \delta \right\}.$$

(a) Determine $D(X\|\hat{X})$ if $P_X(1) = \frac{1}{3} = 1 - P_X(0)$ and $P_{\hat{X}}(1) = \frac{2}{3} = 1 - P_{\hat{X}}(0)$.

(b) Continue from (a). List the elements in $\mathcal{A}_2(0.5)$.

(c) Show that for any sequence x^n in $\mathcal{A}_n(\delta)$,

$$P_{X^n}(x^n)2^{-n(D(P_X\|P_{\hat{X}})-\delta)} > P_{\hat{X}^n}(x^n) > P_{X^n}(x^n)2^{-n(D(P_X\|P_{\hat{X}})+\delta)}.$$

(d) Show that

$$P_{\hat{X}^n}(\mathcal{A}_n(\delta)) \le 2^{-n(D(P_X\|P_{\hat{X}})-\delta)}P_{X^n}(\mathcal{A}_n(\delta)).$$

Hint: Use (c).

(e) Show that for any $\mathcal{B}_n \in \mathcal{X}^n$,

$$P_{\hat{X}^n}(\mathcal{B}_n) \quad \ge \quad 2^{-n(D(P_X\|P_{\hat{X}})+\delta)}P_{X^n}(\mathcal{B}_n \cap \mathcal{A}_n(\delta)).$$

Hint: Use (c).

(f) Show that as long as

$$\lim_{n\to\infty} P_{X^n}(\mathcal{A}_n^c(\delta)) = 0 \text{ for arbitrary } \delta > 0,$$

we obtain

$$\lim_{n\to\infty} -\frac{1}{n}\log_2\left(\min_{\mathcal{B}_n\in\mathcal{X}^n : P_{X^n}(\mathcal{B}_n^c)\le\epsilon} P_{\hat{X}^n}(\mathcal{B}_n)\right) = D(P_X\|P_{\hat{X}}).$$

Hint: Use (d) and (e).

Solution.

(a)

$$D(X\|\hat{X}) = \frac{1}{3}\log_2\frac{1/3}{2/3} + \frac{2}{3}\log_2\frac{2/3}{1/3} = \frac{1}{3}$$

(b) Since

$$\frac{1}{2}\log_2\frac{P_{X^2}(x^2)}{P_{\hat{X}^2}(x^2)} = \begin{cases} \frac{1}{2}\log_2\frac{4/9}{1/9}, & x^2 = 00 \\ \frac{1}{2}\log_2\frac{2/9}{2/9}, & x^2 = 01 \text{ or } 10 \\ \frac{1}{2}\log_2\frac{1/9}{4/9}, & x^2 = 11 \end{cases} = \begin{cases} 1, & x^2 = 00 \\ 0, & x^2 = 01 \text{ or } 10 \\ -1, & x^2 = 11 \end{cases}$$

we obtain $\mathcal{A}_2(0.1) = \{01, 10\}$.

(c)

$$
\begin{aligned}
& \left| \frac{1}{n} \log_2 \frac{P_{X^n}(x^n)}{P_{\hat{X}^n}(x^n)} - D(P_X \| P_{\hat{X}}) \right| < \delta \\
& \quad \Leftrightarrow \quad -\delta < \frac{1}{n} \log_2 \frac{P_{X^n}(x^n)}{P_{\hat{X}^n}(x^n)} - D(P_X \| P_{\hat{X}}) < \delta \\
& \quad \Leftrightarrow \quad D(P_X \| P_{\hat{X}}) - \delta < \frac{1}{n} \log_2 \frac{P_{X^n}(x^n)}{P_{\hat{X}^n}(x^n)} < D(P_X \| P_{\hat{X}}) + \delta \\
& \quad \Leftrightarrow \quad n(D(P_X \| P_{\hat{X}}) - \delta) < \log_2 \frac{P_{X^n}(x^n)}{P_{\hat{X}^n}(x^n)} < n(D(P_X \| P_{\hat{X}}) + \delta) \\
& \quad \Leftrightarrow \quad 2^{n(D(P_X \| P_{\hat{X}}) - \delta)} < \frac{P_{X^n}(x^n)}{P_{\hat{X}^n}(x^n)} < 2^{n(D(P_X \| P_{\hat{X}}) + \delta)} \\
& \quad \Leftrightarrow \quad 2^{-n(D(P_X \| P_{\hat{X}}) - \delta)} > \frac{P_{\hat{X}^n}(x^n)}{P_{X^n}(x^n)} > 2^{-n(D(P_X \| P_{\hat{X}}) + \delta)} \\
& \quad \Leftrightarrow \quad P_{X^n}(x^n) 2^{-n(D(P_X \| P_{\hat{X}}) - \delta)} > P_{\hat{X}^n}(x^n) > P_{X^n}(x^n) 2^{-n(D(P_X \| P_{\hat{X}}) + \delta)}
\end{aligned}
$$

(d)

$$
\begin{aligned}
P_{\hat{X}^n}(\mathcal{A}_n(\delta)) &= \sum_{x^n \in \mathcal{A}_n(\delta)} P_{\hat{X}^n}(x^n) \\
&\leq \sum_{x^n \in \mathcal{A}_n(\delta)} P_{X^n}(x^n) 2^{-n(D(P_X \| P_{\hat{X}}) - \delta)} \\
&= 2^{-n(D(P_X \| P_{\hat{X}}) - \delta)} \sum_{x^n \in \mathcal{A}_n(\delta)} P_{X^n}(x^n) \\
&= 2^{-n(D(P_X \| P_{\hat{X}}) - \delta)} P_{X^n}(\mathcal{A}_n(\delta))
\end{aligned}
$$

(e)

$$
\begin{aligned}
P_{\hat{X}^n}(\mathcal{B}_n) &\geq P_{\hat{X}^n}(\mathcal{B}_n \cap \mathcal{A}_n(\delta)) \\
&= \sum_{x^n \in \mathcal{B}_n \cap \mathcal{A}_n(\delta)} P_{\hat{X}^n}(x^n) \\
&\geq \sum_{x^n \in \mathcal{B}_n \cap \mathcal{A}_n(\delta)} P_{X^n}(x^n) 2^{-n(D(P_X \| P_{\hat{X}}) + \delta)} \\
&= 2^{-n(D(P_X \| P_{\hat{X}}) + \delta)} P_{X^n}(\mathcal{B}_n \cap \mathcal{A}_n(\delta)).
\end{aligned}
$$

(f) As $\lim_{n\to\infty} P_{X^n}(\mathcal{A}_n^c) = 0$, we infer that for **sufficiently large** n,

$$\begin{aligned}
&-\frac{1}{n}\log_2\left(\min_{\mathcal{B}_n\in\mathcal{X}^n: P_{X^n}(\mathcal{B}_n^c)\le\epsilon} P_{\hat{X}^n}(\mathcal{B}_n)\right)\\
&\quad\ge\ -\frac{1}{n}\log_2\left(P_{\hat{X}^n}(\mathcal{A}_n)\right)\\
&\quad\ge\ -\frac{1}{n}\log_2\left(2^{-n(D(P_X\|P_{\hat{X}})-\delta)}\right)\\
&\quad=\ D(P_X\|P_{\hat{X}})-\delta,
\end{aligned}$$

where

$$\underbrace{P_{\hat{X}^n}(\mathcal{A}_n(\delta))\ \le 2^{-n(D(P_X\|P_{\hat{X}})-\delta)}P_{X^n}(\mathcal{A}_n(\delta))}_{\text{(d)}} \le 2^{-n(D(P_X\|P_{\hat{X}})-\delta)}.$$

On the other hand, for **every** n, we infer from (e) that

$$\begin{aligned}
&-\frac{1}{n}\log_2\left(\min_{\mathcal{B}_n\in\mathcal{X}^n: P_{X^n}(\mathcal{B}_n^c)\le\epsilon} P_{\hat{X}^n}(\mathcal{B}_n)\right)\\
&\le\ -\frac{1}{n}\log_2\left(\min_{\mathcal{B}_n\in\mathcal{X}^n: P_{X^n}(\mathcal{B}_n^c)\le\epsilon}\left[2^{-n(D(P_X\|P_{\hat{X}})+\delta)}P_{X^n}(\mathcal{B}_n\cap\mathcal{A}_n(\delta))\right]\right)\\
&=\ -\frac{1}{n}\log_2\left(2^{-n(D(P_X\|P_{\hat{X}})+\delta)}\min_{\mathcal{B}_n\in\mathcal{X}^n: P_{X^n}(\mathcal{B}_n^c)\le\epsilon}\left[P_{X^n}(\mathcal{B}_n\cap\mathcal{A}_n(\delta))\right]\right)\\
&=\ D(P_X\|P_{\hat{X}})+\delta-\frac{1}{n}\log_2\left(\min_{\mathcal{B}_n\in\mathcal{X}^n: P_{X^n}(\mathcal{B}_n^c)\le\epsilon}\left[P_{X^n}(\mathcal{B}_n\cap\mathcal{A}_n(\delta))\right]\right)\\
&\le\ D(P_X\|P_{\hat{X}})+\delta-\frac{1}{n}\log_2\left(\min_{\mathcal{B}_n\in\mathcal{X}^n: P_{X^n}(\mathcal{B}_n^c)\le\epsilon}\left[1-P_{X^n}(\mathcal{B}_n^c)-P_{X^n}(\mathcal{A}_n^c(\delta))\right]\right)\\
&=\ D(P_X\|P_{\hat{X}})+\delta-\frac{1}{n}\log_2\left(1-\max_{\mathcal{B}_n\in\mathcal{X}^n: P_{X^n}(\mathcal{B}_n^c)\le\epsilon}P_{X^n}(\mathcal{B}_n^c)-P_{X^n}(\mathcal{A}_n^c(\delta))\right)\\
&=\ D(P_X\|P_{\hat{X}})+\delta-\frac{1}{n}\log_2\left(1-\epsilon-P_{X^n}(\mathcal{A}_n^c(\delta))\right).
\end{aligned}$$

We then conclude

$$\begin{aligned}
D(P_X\|P_{\hat{X}})-\delta\ &\le\ \liminf_{n\to\infty}-\frac{1}{n}\log_2\left(\min_{\mathcal{B}_n\in\mathcal{X}^n: P_{X^n}(\mathcal{B}_n^c)\le\epsilon} P_{\hat{X}^n}(\mathcal{B}_n)\right)\\
&\le\ \limsup_{n\to\infty}-\frac{1}{n}\log_2\left(\min_{\mathcal{B}_n\in\mathcal{X}^n: P_{X^n}(\mathcal{B}_n^c)\le\epsilon} P_{\hat{X}^n}(\mathcal{B}_n)\right)\\
&\le\ D(P_X\|P_{\hat{X}})+\delta.
\end{aligned}$$

Since δ can be made arbitrarily small, the desired result holds.