香港中文大學
The Chinese University of Hong Kong

2.8 Fano's Inequality

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|,$$

where $|\mathcal{X}|$ denotes the size of the alphabet $\mathcal{X}$. This upper bound is tight if and only if X is distributed uniformly on $\mathcal{X}$.

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

Theorem 2.43 For any random variable X,

$$H(X) \le \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

Theorem 2.43 For any random variable X,

$$H(X) \le \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

Theorem 2.43 For any random variable X,

$$H(X) \le \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\log |\mathcal{X}| - H(X)$$

Theorem 2.43 For any random variable X,

$$H(X) \le \log|\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned} &\log|\mathcal{X}| - H(X) \\ &\quad = \quad -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log|\mathcal{X}|, \quad (1)$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\underline{\log|\mathcal{X}|} - H(X) \\
&\quad = \underline{-\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|}} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x)
\end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned} &\log |\mathcal{X}| - H(X) \\ &\quad = \quad -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned} &\log |\mathcal{X}| \underline{- H(X)} \\ &\quad = \quad -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \underline{\sum_{x \in \mathcal{S}_X} p(x) \log p(x)} \end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned} &\log |\mathcal{X}| - H(X) \\ &\quad = \quad -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned} &\log |\mathcal{X}| - H(X) \\ &\quad = \quad - \sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{\underline{|\mathcal{X}|}} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = \quad -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = \quad -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x)
\end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x)
\end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \le \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \underline{\log u(x)} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x)
\end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \underline{\log u(x)} + \sum_{x \in \mathcal{S}_X} p(x) \underline{\log p(x)}
\end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \underline{\log u(x)} + \sum_{x \in \mathcal{S}_X} p(x) \underline{\log p(x)} \\
&\quad = \sum_{x \in \mathcal{S}_X} p(x) \underline{\log \frac{p(x)}{u(x)}}
\end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \le \log|\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log|\mathcal{X}| - H(X) \\
&= -\sum_{x\in\mathcal{S}_X} p(x)\log\frac{1}{|\mathcal{X}|} + \sum_{x\in\mathcal{S}_X} p(x)\log p(x) \\
&= -\sum_{x\in\mathcal{S}_X} p(x)\,\underline{\log u(x)} + \sum_{x\in\mathcal{S}_X} p(x)\,\underline{\log p(x)} \\
&= \sum_{x\in\mathcal{S}_X} p(x)\,\underline{\log\frac{p(x)}{u(x)}}
\end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \underline{\log u(x)} + \sum_{x \in \mathcal{S}_X} p(x) \underline{\log p(x)} \\
&\quad = \sum_{x \in \mathcal{S}_X} p(x) \underline{\log \frac{p(x)}{u(x)}}
\end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \le \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = \sum_{x \in \mathcal{S}_X} p(x) \log \frac{p(x)}{u(x)} \\
&\quad = D(p \| u)
\end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = \sum_{x \in \mathcal{S}_X} p(x) \log \frac{p(x)}{u(x)} \\
&\quad = D(p \| u) \\
&\quad \geq 0,
\end{aligned}$$

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = \sum_{x \in \mathcal{S}_X} p(x) \log \frac{p(x)}{u(x)} \\
&\quad = D(p \| u) \\
&\quad \geq 0,
\end{aligned}$$

proving (1).

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = \quad -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = \quad -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = \quad \sum_{x \in \mathcal{S}_X} p(x) \log \frac{p(x)}{u(x)} \\
&\quad = \quad D(p \| u) \\
&\quad \geq \quad 0,
\end{aligned}$$

proving (1).

3. This upper bound is tight if and if $p = u$, i.e., X is uniformly distributed on $|\mathcal{X}|$.

Theorem 2.43 For any random variable X,

$$H(X) \leq \log|\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log|\mathcal{X}| - H(X) \\
&= -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&= -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&= \sum_{x \in \mathcal{S}_X} p(x) \log \frac{p(x)}{u(x)} \\
&= D(p\|u) \\
&\geq 0,
\end{aligned}$$

proving (1).

3. This upper bound is tight if and if $p = u$, i.e., X is uniformly distributed on $|\mathcal{X}|$.

Theorem 2.43 For any random variable X,

$$H(X) \le \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = \sum_{x \in \mathcal{S}_X} p(x) \log \frac{p(x)}{u(x)} \\
&\quad = D(p \| u) \\
&\quad \ge 0,
\end{aligned}$$

proving (1).

3. This upper bound is tight if and if $p = u$, i.e., X is uniformly distributed on $|\mathcal{X}|$.

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned} & \log |\mathcal{X}| - H(X) \\ &= -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\ &= -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\ &= \sum_{x \in \mathcal{S}_X} p(x) \log \frac{p(x)}{u(x)} \\ &= D(p \| u) \\ &\geq 0, \end{aligned}$$

proving (1).

3. This upper bound is tight if and if $p = u$, i.e., X is uniformly distributed on $|\mathcal{X}|$.

Remark Theorem 2.43 has the following implications.

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&\quad = \sum_{x \in \mathcal{S}_X} p(x) \log \frac{p(x)}{u(x)} \\
&\quad = D(p \| u) \\
&\quad \geq 0,
\end{aligned}$$

proving (1).

3. This upper bound is tight if and if $p = u$, i.e., X is uniformly distributed on $|\mathcal{X}|$.

Remark Theorem 2.43 has the following implications.

1. For a random variable X, if the alphabet is finite, then

$$H(X) \leq \log |\mathcal{X}| < \infty,$$

i.e., $H(X)$ is finite.

Theorem 2.43 For any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned} &\log |\mathcal{X}| - H(X) \\ &= -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\ &= -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\ &= \sum_{x \in \mathcal{S}_X} p(x) \log \frac{p(x)}{u(x)} \\ &= D(p\|u) \\ &\geq 0, \end{aligned}$$

proving (1).

3. This upper bound is tight if and if $p = u$, i.e., X is uniformly distributed on $|\mathcal{X}|$.

Remark Theorem 2.43 has the following implications.

1. For a random variable X, if the alphabet is finite, then

$$H(X) \leq \log |\mathcal{X}| < \infty,$$

i.e., $H(X)$ is finite.

2. The entropy of a random variable may take any nonnegative real value, i.e., for any $a \geq 0$, we can find a random variable X such that $H(X) = a$. See Corollary 2.44.

Theorem 2.43 For any random variable X,

$$H(X) \le \log |\mathcal{X}|, \tag{1}$$

with equality if and only if X is distributed uniformly on $\mathcal{X}$.

Proof

1. Let u be the uniform distribution on $\mathcal{X}$, i.e.,

$$u(x) = \frac{1}{|\mathcal{X}|} \quad \text{for all } x \in \mathcal{X}.$$

2. Then

$$\begin{aligned}
&\log |\mathcal{X}| - H(X) \\
&= -\sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{|\mathcal{X}|} + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&= -\sum_{x \in \mathcal{S}_X} p(x) \log u(x) + \sum_{x \in \mathcal{S}_X} p(x) \log p(x) \\
&= \sum_{x \in \mathcal{S}_X} p(x) \log \frac{p(x)}{u(x)} \\
&= D(p \| u) \\
&\ge 0,
\end{aligned}$$

proving (1).

3. This upper bound is tight if and if $p = u$, i.e., X is uniformly distributed on $|\mathcal{X}|$.

Remark Theorem 2.43 has the following implications.

1. For a random variable X, if the alphabet is finite, then

$$H(X) \le \log |\mathcal{X}| < \infty,$$

   i.e., $H(X)$ is finite.

2. The entropy of a random variable may take any nonnegative real value, i.e., for any $a \ge 0$, we can find a random variable X such that $H(X) = a$. See Corollary 2.44.

Remark For a random variable X, if the alphabet is infinite, then $H(X)$ can be finite (Example 2.45) or infinite (Example 2.46).

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$.

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$H_2(X) = -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$H_2(X) = -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$H_2(X) = -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$H_2(X) = -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$H_2(X) = -\sum_{i=1}^{\infty} 2^{-i} \underline{\log 2^{-i}}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \underline{\log 2^{-i}} \\ &= -\sum_{i=1}^{\infty} 2^{-i} \underline{(-i)} \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j.

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j.

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i 2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j.

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j.

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{(i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i}\right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j.

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i 2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j.

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \underline{\Pr\{Y = (i, j)\}} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \underline{\Pr\{Y = (i, j)\}} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) \underline{2^{-2^i}} = 1.$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i 2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\underline{2^{2^i}/2^i}} \underline{\Pr\{Y = (i, j)\}} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{\underline{2^i}} \right) \underline{2^{-2^i}} = 1.$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) \cancel{2^{-2^i}} = 1.$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{\cancel{2^{2^i}}}{2^i} \right) \cancel{2^{-2^i}} = 1.$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$H_2(Y) = -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$H_2(Y) = -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i 2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$H_2(Y) = -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$H_2(Y) = -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$H_2(Y) = -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \underline{\log_2 2^{-2^i}}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \underline{\log_2 2^{-2^i}} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} (\underline{-2^i}) \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} (-2^i) \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \underline{2^{-2^i}(-2^i)} \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \underline{2^{-2^i}(-2^i)} \\ &= \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) \underline{2^{-2^i} \cdot 2^i} \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \underline{2^{-2^i}(-2^i)} \\ &= \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) \underline{2^{-2^i} \cdot 2^i} \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned}
H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\
&= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\
&= \sum_{i=1}^{\infty} i 2^{-i}.
\end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned}
H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\
&= -\sum_{i=1}^{\infty} \sum_{j=1}^{\underline{2^{2^i}/2^i}} \underline{2^{-2^i}(-2^i)} \\
&= \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{\underline{2^i}} \right) \underline{2^{-2^i} \cdot 2^i}
\end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} (-2^i) \\ &= \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} \cdot 2^i \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} (-2^i) \\ &= \sum_{i=1}^{\infty} \left(\frac{\cancel{2^{2^i}}}{2^i} \right) 2^{-2^i} \cdot 2^i \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} (-2^i) \\ &= \sum_{i=1}^{\infty} \left(\frac{\cancel{2^{2^i}}}{2^i} \right) \cancel{2^{-2^i}} \cdot 2^i \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} (-2^i) \\ &= \sum_{i=1}^{\infty} \left(\frac{\cancel{2^{2^i}}}{\cancel{2^i}} \right) \cancel{2^{-2^i}} \cdot 2^i \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i 2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} (-2^i) \\ &= \sum_{i=1}^{\infty} \left(\frac{\cancel{2^{2^i}}}{\cancel{2^i}} \right) \cancel{2^{-2^i}} \cdot \cancel{2^i} \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i 2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \le i < \infty \text{ and } 1 \le j \le \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} (-2^i) \\ &= \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} \cdot 2^i \\ &= \sum_{i=1}^{\infty} 1, \end{aligned}$$

Example 2.45

Let X be a random variable such that

$$\Pr\{X = i\} = 2^{-i},$$

$i = 1, 2, \cdots$. Then

$$\begin{aligned} H_2(X) &= -\sum_{i=1}^{\infty} 2^{-i} \log 2^{-i} \\ &= -\sum_{i=1}^{\infty} 2^{-i}(-i) \\ &= \sum_{i=1}^{\infty} i2^{-i}. \end{aligned}$$

This is the expectation of a truncated geometric distribution, which is equal to 2.

Example 2.46

Let Y be a random variable which takes values in the subset of pairs of integers

$$\left\{ (i, j) : 1 \leq i < \infty \text{ and } 1 \leq j \leq \frac{2^{2^i}}{2^i} \right\}$$

such that

$$\Pr\{Y = (i, j)\} = 2^{-2^i}$$

for all i and j. First, we check that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} \Pr\{Y = (i, j)\} = \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} = 1.$$

Then

$$\begin{aligned} H_2(Y) &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i} \log_2 2^{-2^i} \\ &= -\sum_{i=1}^{\infty} \sum_{j=1}^{2^{2^i}/2^i} 2^{-2^i}(-2^i) \\ &= \sum_{i=1}^{\infty} \left(\frac{2^{2^i}}{2^i} \right) 2^{-2^i} \cdot 2^i \\ &= \sum_{i=1}^{\infty} 1, \end{aligned}$$

which does not converge.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Intuition

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Intuition

- Suppose $\hat{X}$ is an estimate on X.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Intuition

- Suppose $\hat{X}$ is an estimate on X.
- If the error probability P_e is small, then $H(X|\hat{X})$ should also be small.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Intuition

- Suppose $\hat{X}$ is an estimate on X.
- If the error probability P_e is small, then $H(X|\hat{X})$ should also be small.
- This is captured by Fano's inequality.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Intuition

- Suppose $\hat{X}$ is an estimate on X.
- If the error probability P_e is small, then $H(X|\hat{X})$ should also be small.
- This is captured by Fano's inequality.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Intuition

- Suppose $\hat{X}$ is an estimate on X.
- If the error probability P_e is small, then $H(X|\hat{X})$ should also be small.
- This is captured by Fano's inequality.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Intuition

- Suppose $\hat{X}$ is an estimate on X.
- If the error probability P_e is small, then $H(X|\hat{X})$ should also be small.
- This is captured by Fano's inequality.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \ne \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \ne \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \ne \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ \underline{1 \quad \text{if } X \neq \hat{X}}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{\underline{Y = 1}\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$H(X|\hat{X})$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \ne \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \ne \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \ne \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \ne \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \ne \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \ne \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \ne \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \ne \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \ne \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

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

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
&H(X|\hat{X}) \\
&\quad= H(X|\hat{X}) + H(Y|X, \hat{X}) \\
&\quad= H(X, Y|\hat{X}) \\
&\quad= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
&\quad\leq H(Y) + \underline{H(X|\hat{X}, Y)} \\
&\quad= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
&\qquad \Big[\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0) \\
&\qquad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big].
\end{aligned}$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
& H(X|\hat{X}) \\
&= H(X|\hat{X}) + H(Y|X, \hat{X}) \\
&= H(X, Y|\hat{X}) \\
&= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
&\leq H(Y) + \underline{H(X|\hat{X}, Y)} \\
&= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
& \quad \Big[\Pr\{\underline{\hat{X} = \hat{x}, Y = 0}\} H(X|\hat{X} = \hat{x}, Y = 0) \\
& \quad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big].
\end{aligned}$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
& H(X|\hat{X}) \\
= \quad & H(X|\hat{X}) + H(Y|X, \hat{X}) \\
= \quad & H(X, Y|\hat{X}) \\
= \quad & H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
\le \quad & H(Y) + \underline{H(X|\hat{X}, Y)} \\
= \quad & H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
& \Big[\Pr\{\underline{\hat{X} = \hat{x}, Y = 0}\} H(X|\underline{\hat{X} = \hat{x}, Y = 0}) \\
& + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big].
\end{aligned}$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \ne \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \ne \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \ne \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
&H(X|\hat{X}) \\
&\quad = \quad H(X|\hat{X}) + H(Y|X, \hat{X}) \\
&\quad = \quad H(X, Y|\hat{X}) \\
&\quad = \quad H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
&\quad \le \quad H(Y) + \underline{H(X|\hat{X}, Y)} \\
&\quad = \quad H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
&\qquad\qquad \Big[\Pr\{\underline{\hat{X} = \hat{x}, Y = 0}\} H(X|\underline{\hat{X} = \hat{x}, Y = 0}) \\
&\qquad\qquad + \Pr\{\underline{\hat{X} = \hat{x}, Y = 1}\} H(X|\hat{X} = \hat{x}, Y = 1)\Big].
\end{aligned}$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
&H(X|\hat{X}) \\
&\quad= \quad H(X|\hat{X}) + H(Y|X, \hat{X}) \\
&\quad= \quad H(X, Y|\hat{X}) \\
&\quad= \quad H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
&\quad\leq \quad H(Y) + \underline{H(X|\hat{X}, Y)} \\
&\quad= \quad H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
&\qquad\qquad \Big[\Pr\{\underline{\hat{X} = \hat{x}, Y = 0}\} H(X|\underline{\hat{X} = \hat{x}, Y = 0}) \\
&\qquad\qquad + \Pr\{\underline{\hat{X} = \hat{x}, Y = 1}\} H(X|\underline{\hat{X} = \hat{x}, Y = 1})\Big].
\end{aligned}$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
&H(X|\hat{X}) \\
&\quad = H(X|\hat{X}) + H(Y|X, \hat{X}) \\
&\quad = H(X, Y|\hat{X}) \\
&\quad = H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
&\quad \le H(Y) + H(X|\hat{X}, Y) \\
&\quad = H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
&\qquad \Big[\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0) \\
&\qquad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big].
\end{aligned}$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ = \quad & H(X|\hat{X}) + H(Y|X, \hat{X}) \\ = \quad & H(X, Y|\hat{X}) \\ = \quad & H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ \leq \quad & H(Y) + H(X|\hat{X}, Y) \\ = \quad & H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ & \Big[\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0) \\ & + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ &= H(X|\hat{X}) + H(Y|X, \hat{X}) \\ &= H(X, Y|\hat{X}) \\ &= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ &\le H(Y) + H(X|\hat{X}, Y) \\ &= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ & \quad \Big[\Pr\{\underline{\hat{X} = \hat{x}, Y = 0}\} H(X|\hat{X} = \hat{x}, Y = 0) \\ & \quad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\underline{\hat{X} = \hat{x} \text{ and } Y = 0,}$ X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \ne \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \ne \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \ne \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ = \quad & H(X|\hat{X}) + H(Y|X,\hat{X}) \\ = \quad & H(X, Y|\hat{X}) \\ = \quad & H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ \le \quad & H(Y) + H(X|\hat{X}, Y) \\ = \quad & H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ & \Big[\Pr\{\underline{\hat{X} = \hat{x}, Y = 0}\} H(X|\hat{X} = \hat{x}, Y = 0) \\ & + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\underline{\hat{X} = \hat{x} \text{ and } Y = 0}$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
&H(X|\hat{X}) \\
&= H(X|\hat{X}) + H(Y|X, \hat{X}) \\
&= H(X, Y|\hat{X}) \\
&= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
&\leq H(Y) + H(X|\hat{X}, Y) \\
&= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
&\quad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\
&\quad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big].
\end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
&H(X|\hat{X}) \\
&= H(X|\hat{X}) + H(Y|X, \hat{X}) \\
&= H(X, Y|\hat{X}) \\
&= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
&\leq H(Y) + H(X|\hat{X}, Y) \\
&= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
&\quad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\
&\quad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big].
\end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ = \quad & H(X|\hat{X}) + H(Y|X, \hat{X}) \\ = \quad & H(X, Y|\hat{X}) \\ = \quad & H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ \le \quad & H(Y) + H(X|\hat{X}, Y) \\ = \quad & H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ & \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ & + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ = \quad & H(X|\hat{X}) + H(Y|X, \hat{X}) \\ = \quad & H(X, Y|\hat{X}) \\ = \quad & H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ \leq \quad & H(Y) + H(X|\hat{X}, Y) \\ = \quad & H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ & \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ & + \Pr\{\underline{\hat{X} = \hat{x}, Y = 1}\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\underline{\hat{X} = \hat{x} \text{ and } Y = 1}$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ = \quad & H(X|\hat{X}) + H(Y|X, \hat{X}) \\ = \quad & H(X, Y|\hat{X}) \\ = \quad & H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ \leq \quad & H(Y) + H(X|\hat{X}, Y) \\ = \quad & H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ & \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ & + \Pr\{\underline{\hat{X} = \hat{x}, Y = 1}\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\underline{\hat{X} = \hat{x} \text{ and } Y = 1}$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

Theorem 2.43 $H(X) \leq \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ & \quad = H(X|\hat{X}) + H(Y|X, \hat{X}) \\ & \quad = H(X, Y|\hat{X}) \\ & \quad = H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ & \quad \leq H(Y) + H(X|\hat{X}, Y) \\ & \quad = H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ & \qquad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ & \qquad + \Pr\{\underline{\hat{X} = \hat{x}, Y = 1}\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\underline{\hat{X} = \hat{x} \text{ and } Y = 1}$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \leq \log(|\mathcal{X}| - 1),$$

Theorem 2.43 $H(X) \leq \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ &= H(X|\hat{X}) + H(Y|X, \hat{X}) \\ &= H(X, Y|\hat{X}) \\ &= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ &\leq H(Y) + H(X|\hat{X}, Y) \\ &= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ &\quad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ &\quad + \Pr\{\underline{\hat{X} = \hat{x}, Y = 1}\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\underline{\hat{X} = \hat{x} \text{ and } Y = 1}$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \leq \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

$\boxed{\textbf{Theorem 2.43} \quad H(X) \leq \log|\mathcal{X}|.}$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ &= H(X|\hat{X}) + H(Y|X, \hat{X}) \\ &= H(X, Y|\hat{X}) \\ &= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ &\le H(Y) + H(X|\hat{X}, Y) \\ &= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ &\quad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ &\quad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \le \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$\boxed{\textbf{Theorem 2.43} \quad H(X) \le \log |\mathcal{X}|.}$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \ne \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \ne \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \ne \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
&H(X|\hat{X}) \\
&= H(X|\hat{X}) + H(Y|X, \hat{X}) \\
&= H(X, Y|\hat{X}) \\
&= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
&\le H(Y) + H(X|\hat{X}, Y) \\
&= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
&\quad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\
&\quad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big].
\end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \ne \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \le \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$H(X|\hat{X})$$

Theorem 2.43 $H(X) \le \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \ne \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \ne \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \ne \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ = \quad & H(X|\hat{X}) + H(Y|X, \hat{X}) \\ = \quad & H(X, Y|\hat{X}) \\ = \quad & H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ \le \quad & H(Y) + H(X|\hat{X}, Y) \\ = \quad & H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ & \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ & + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \ne \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \le \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned} & H(X|\hat{X}) \\ \le \quad & h_b(P_e) + \left(\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}\right) \log(|\mathcal{X}| - 1) \end{aligned}$$

$$\boxed{\textbf{Theorem 2.43} \quad H(X) \le \log |\mathcal{X}|.}$$

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \ne \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \ne \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \ne \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
&H(X|\hat{X}) \\
&\quad = H(X|\hat{X}) + H(Y|X, \hat{X}) \\
&\quad = H(X, Y|\hat{X}) \\
&\quad = H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
&\quad \le H(Y) + H(X|\hat{X}, Y) \\
&\quad = H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
&\qquad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\
&\qquad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big].
\end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \ne \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \le \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned}
&H(X|\hat{X}) \\
&\quad \le \underline{h_b(P_e)} + \left(\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}\right) \log(|\mathcal{X}| - 1)
\end{aligned}$$

Theorem 2.43 $H(X) \le \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \ne \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \ne \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \ne \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ & = H(X|\hat{X}) + H(Y|X, \hat{X}) \\ & = H(X, Y|\hat{X}) \\ & = H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ & \le H(Y) + H(X|\hat{X}, Y) \\ & = \underline{H(Y)} + \sum_{\hat{x} \in \mathcal{X}} \\ & \quad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ & \quad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \ne \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \le \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned} & H(X|\hat{X}) \\ & \le \underline{h_b(P_e)} + \left(\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}\right) \log(|\mathcal{X}| - 1) \end{aligned}$$

Theorem 2.43 $H(X) \le \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ = \;& H(X|\hat{X}) + H(Y|X, \hat{X}) \\ = \;& H(X, Y|\hat{X}) \\ = \;& H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ \le \;& H(Y) + H(X|\hat{X}, Y) \\ = \;& H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ & \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ & + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \le \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned} & H(X|\hat{X}) \\ \le \;& h_b(P_e) + \left(\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}\right) \log(|\mathcal{X}| - 1) \end{aligned}$$

Theorem 2.43 $H(X) \le \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} &H(X|\hat{X}) \\ &= H(X|\hat{X}) + H(Y|X, \hat{X}) \\ &= H(X, Y|\hat{X}) \\ &= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ &\le H(Y) + H(X|\hat{X}, Y) \\ &= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ &\qquad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \le \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned} &H(X|\hat{X}) \\ &\le h_b(P_e) + \left(\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}\right) \underline{\log(|\mathcal{X}| - 1)} \end{aligned}$$

Theorem 2.43 $H(X) \le \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ = \quad & H(X|\hat{X}) + H(Y|X, \hat{X}) \\ = \quad & H(X, Y|\hat{X}) \\ = \quad & H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ \leq \quad & H(Y) + H(X|\hat{X}, Y) \\ = \quad & H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ & \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ & + \Pr\{\hat{X} = \hat{x}, Y = 1\} \underline{H(X|\hat{X} = \hat{x}, Y = 1)}\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \leq \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned} & H(X|\hat{X}) \\ \leq \quad & h_b(P_e) + \left(\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}\right) \underline{\log(|\mathcal{X}| - 1)} \end{aligned}$$

Theorem 2.43 $H(X) \leq \log|\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
&H(X|\hat{X}) \\
&\quad = H(X|\hat{X}) + H(Y|X, \hat{X}) \\
&\quad = H(X, Y|\hat{X}) \\
&\quad = H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
&\quad \leq H(Y) + H(X|\hat{X}, Y) \\
&\quad = H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
&\qquad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\
&\qquad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big].
\end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \leq \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned}
&H(X|\hat{X}) \\
&\quad \leq h_b(P_e) + \left(\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}\right) \log(|\mathcal{X}| - 1)
\end{aligned}$$

Theorem 2.43 $H(X) \leq \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \ne \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \ne \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \ne \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ = \quad & H(X|\hat{X}) + H(Y|X, \hat{X}) \\ = \quad & H(X, Y|\hat{X}) \\ = \quad & H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ \le \quad & H(Y) + H(X|\hat{X}, Y) \\ = \quad & H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ & \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ & + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \ne \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \le \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned} & H(X|\hat{X}) \\ \le \quad & h_b(P_e) + \left(\underline{\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}}\right) \log(|\mathcal{X}| - 1) \end{aligned}$$

Theorem 2.43 $H(X) \le \log|\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ &= H(X|\hat{X}) + H(Y|X, \hat{X}) \\ &= H(X, Y|\hat{X}) \\ &= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ &\le H(Y) + H(X|\hat{X}, Y) \\ &= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ &\quad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ &\quad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \le \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned} & H(X|\hat{X}) \\ &\le h_b(P_e) + \left(\underline{\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}}\right) \log(|\mathcal{X}| - 1) \\ &= h_b(P_e) + \underline{\Pr\{Y = 1\}} \log(|\mathcal{X}| - 1) \end{aligned}$$

Theorem 2.43 $H(X) \le \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned}
&H(X|\hat{X}) \\
&= H(X|\hat{X}) + H(Y|X, \hat{X}) \\
&= H(X, Y|\hat{X}) \\
&= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\
&\leq H(Y) + H(X|\hat{X}, Y) \\
&= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\
&\quad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\
&\quad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big].
\end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \leq \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned}
&H(X|\hat{X}) \\
&\leq h_b(P_e) + \left(\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}\right) \log(|\mathcal{X}| - 1) \\
&= h_b(P_e) + \underline{\Pr\{Y = 1\}} \log(|\mathcal{X}| - 1)
\end{aligned}$$

Theorem 2.43 $H(X) \leq \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} & H(X|\hat{X}) \\ &= H(X|\hat{X}) + H(Y|X, \hat{X}) \\ &= H(X, Y|\hat{X}) \\ &= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ &\le H(Y) + H(X|\hat{X}, Y) \\ &= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ &\quad \Big[\cancel{\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0)} \\ &\quad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \le \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned} & H(X|\hat{X}) \\ &\le h_b(P_e) + \left(\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}\right) \log(|\mathcal{X}| - 1) \\ &= h_b(P_e) + \underline{\Pr\{Y = 1\}} \log(|\mathcal{X}| - 1) \\ &= h_b(P_e) + \underline{P_e} \log(|\mathcal{X}| - 1), \end{aligned}$$

Theorem 2.43 $H(X) \le \log|\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality)

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \Pr\{X \neq \hat{X}\}$.

Proof

1. Define a random variable

$$Y = \begin{cases} 0 & \text{if } X = \hat{X} \\ 1 & \text{if } X \neq \hat{X}. \end{cases}$$

Y is an indicator of the error event $\{X \neq \hat{X}\}$. Then

$$\Pr\{Y = 1\} = P_e \quad \text{and} \quad H(Y) = h_b(P_e).$$

2. Since Y is a function X and $\hat{X}$,

$$H(Y|X, \hat{X}) = 0.$$

3. Then

$$\begin{aligned} &H(X|\hat{X}) \\ &= H(X|\hat{X}) + H(Y|X, \hat{X}) \\ &= H(X, Y|\hat{X}) \\ &= H(Y|\hat{X}) + H(X|\hat{X}, Y) \\ &\leq H(Y) + H(X|\hat{X}, Y) \\ &= H(Y) + \sum_{\hat{x} \in \mathcal{X}} \\ &\quad \Big[\Pr\{\hat{X} = \hat{x}, Y = 0\} H(X|\hat{X} = \hat{x}, Y = 0) \\ &\quad + \Pr\{\hat{X} = \hat{x}, Y = 1\} H(X|\hat{X} = \hat{x}, Y = 1)\Big]. \end{aligned}$$

4. For the first case, i.e., $\hat{X} = \hat{x}$ and $Y = 0$, X must take the value $\hat{x}$. In other words, X is conditionally deterministic given $\hat{X} = \hat{x}$ and $Y = 0$. Therefore,

$$H(X|\hat{X} = \hat{x}, Y = 0) = 0.$$

5. For the second case, i.e., $\hat{X} = \hat{x}$ and $Y = 1$, X must take a value in the set $\{x \in \mathcal{X} : x \neq \hat{x}\}$ which contains $|\mathcal{X}| - 1$ elements. By Theorem 2.43, we have

$$H(X|\hat{X} = \hat{x}, Y = 1) \leq \log(|\mathcal{X}| - 1),$$

where this upper bound does not depend on $\hat{x}$.

6. Hence,

$$\begin{aligned} &H(X|\hat{X}) \\ &\leq h_b(P_e) + \left(\sum_{\hat{x} \in \mathcal{X}} \Pr\{\hat{X} = \hat{x}, Y = 1\}\right) \log(|\mathcal{X}| - 1) \\ &= h_b(P_e) + \Pr\{Y = 1\} \log(|\mathcal{X}| - 1) \\ &= h_b(P_e) + P_e \log(|\mathcal{X}| - 1), \end{aligned}$$

completing the proof.

Theorem 2.43 $H(X) \leq \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Corollary 2.48 $H(X|\hat{X}) < 1 + P_e \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq \underline{h_b(P_e)} + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Corollary 2.48 $H(X|\hat{X}) < 1 + P_e \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq \underline{h_b(P_e)} + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Corollary 2.48 $H(X|\hat{X}) < \underline{1} + P_e \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Corollary 2.48 $H(X|\hat{X}) < 1 + P_e \log |\mathcal{X}|$.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \le h_b(P_e) + P_e \log(\underline{|\mathcal{X}| - 1}),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Corollary 2.48 $H(X|\hat{X}) < 1 + P_e \log \underline{|\mathcal{X}|}$.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Corollary 2.48 $H(X|\hat{X}) < 1 + P_e \log |\mathcal{X}|$.

Remarks

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Corollary 2.48 $H(X|\hat{X}) < 1 + P_e \log |\mathcal{X}|$.

Remarks

- For finite alphabet, if $P_e \to 0$, then $H(X|\hat{X}) \to 0$.

Theorem 2.47 (Fano's Inequality) Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then

$$H(X|\hat{X}) \leq h_b(P_e) + P_e \log(|\mathcal{X}| - 1),$$

where $P_e = \{X \neq \hat{X}\}$ and h_b is the binary entropy function.

Corollary 2.48 $H(X|\hat{X}) < 1 + P_e \log |\mathcal{X}|$.

Remarks

- For finite alphabet, if $P_e \to 0$, then $H(X|\hat{X}) \to 0$.
- This may NOT hold for countably infinite alphabet (see Example 2.49).