香港中文大學
The Chinese University of Hong Kong

8.4 The Converse

How to Prove the Converse

How to Prove the Converse

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

How to Prove the Converse

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

- Prove that for any achievable rate-distortion pair (R, D), $R \geq R_I(D)$.

How to Prove the Converse

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

- Prove that for any achievable rate-distortion pair (R, D), $R \geq R_I(D)$.

Definition 8.13 The rate-distortion function $R(D)$ is the minimum of all rates R for a given distortion D such that (R, D) is achievable.

How to Prove the Converse

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

- Prove that for any achievable rate-distortion pair (R, D), $R \geq R_I(D)$.
- Fix D and minimize R over all achievable pairs (R, D) to conclude that $R(D) \geq R_I(D)$.

Definition 8.13 The rate-distortion function $R(D)$ is the minimum of all rates R for a given distortion D such that (R, D) is achievable.

The Converse

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$n(R + \epsilon)$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned} & n(R + \epsilon) \\ & \quad \geq \quad \log M \end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned} & n(R + \epsilon) \\ & \quad \geq \quad \log M \\ & \quad \geq \quad H(f(\mathbf{X})) \end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned} n(R + \epsilon) & \\ \geq \quad & \log M \\ \geq \quad & H(f(\mathbf{X})) \\ \geq \quad & H(g(f(\mathbf{X}))) \end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned} & n(R + \epsilon) \\ & \quad \geq \log M \\ & \quad \geq H(f(\mathbf{X})) \\ & \quad \geq H(g(f(\mathbf{X}))) \\ & \quad = H(\hat{\mathbf{X}}) \end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
& n(R + \epsilon) \\
&\quad \geq \quad \log M \\
&\quad \geq \quad H(f(\mathbf{X})) \\
&\quad \geq \quad H(g(f(\mathbf{X}))) \\
&\quad = \quad H(\hat{\mathbf{X}}) \\
&\quad = \quad H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X})
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
& n(R + \epsilon) \\
& \quad \geq \quad \log M \\
& \quad \geq \quad H(f(\mathbf{X})) \\
& \quad \geq \quad H(g(f(\mathbf{X}))) \\
& \quad = \quad H(\hat{\mathbf{X}}) \\
& \quad = \quad H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
& \quad = \quad I(\hat{\mathbf{X}}; \mathbf{X})
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
& n(R+\epsilon) \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}})
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
& n(R + \epsilon) \\
& \quad \geq \quad \log M \\
& \quad \geq \quad H(f(\mathbf{X})) \\
& \quad \geq \quad H(g(f(\mathbf{X}))) \\
& \quad = \quad H(\hat{\mathbf{X}}) \\
& \quad = \quad H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
& \quad = \quad I(\hat{\mathbf{X}}; \mathbf{X}) \\
& \quad = \quad \underline{H(\mathbf{X})} - H(\mathbf{X}|\hat{\mathbf{X}})
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & \underline{H(\mathbf{X})} - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \underline{\sum_{k=1}^{n} H(X_k)} \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1})
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - \underline{H(\mathbf{X}|\hat{\mathbf{X}})} \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1})
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R + \epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - \underline{H(\mathbf{X}|\hat{\mathbf{X}})} \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& \underline{- \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1})}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n}\log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R + \epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} \underline{H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1})}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq &\ \log M \\
\geq &\ H(f(\mathbf{X})) \\
\geq &\ H(g(f(\mathbf{X}))) \\
= &\ H(\hat{\mathbf{X}}) \\
= &\ H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= &\ I(\hat{\mathbf{X}}; \mathbf{X}) \\
= &\ H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= &\ \sum_{k=1}^{n} H(X_k) \\
&\ - \sum_{k=1}^{n} \underline{H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1})} \\
\geq &\ \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} \underline{H(X_k|\hat{X}_k)}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \quad & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k)
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n}\log M \le R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \le \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
&n(R+\epsilon)\\
&\quad\ge \log M\\
&\quad\ge H(f(\mathbf{X}))\\
&\quad\ge H(g(f(\mathbf{X})))\\
&\quad= H(\hat{\mathbf{X}})\\
&\quad= H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X})\\
&\quad= I(\hat{\mathbf{X}};\mathbf{X})\\
&\quad= H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}})\\
&\quad= \sum_{k=1}^{n} H(X_k)\\
&\qquad - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1})\\
&\quad\ge \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k)\\
&\quad= \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)]
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \quad & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \quad & \sum_{k=1}^{n} [\underline{H(X_k) - H(X_k|\hat{X}_k)}]
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n}\log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
&n(R+\epsilon) \\
&\quad \geq \log M \\
&\quad \geq H(f(\mathbf{X})) \\
&\quad \geq H(g(f(\mathbf{X}))) \\
&\quad = H(\hat{\mathbf{X}}) \\
&\quad = H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
&\quad = I(\hat{\mathbf{X}}; \mathbf{X}) \\
&\quad = H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
&\quad = \sum_{k=1}^{n} H(X_k) \\
&\qquad - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
&\quad \geq \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
&\quad = \sum_{k=1}^{n} [\underline{H(X_k) - H(X_k|\hat{X}_k)}] \\
&\quad = \sum_{k=1}^{n} \underline{I(X_k; \hat{X}_k)}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n}\log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
&\geq \log M \\
&\geq H(f(\mathbf{X})) \\
&\geq H(g(f(\mathbf{X}))) \\
&= H(\hat{\mathbf{X}}) \\
&= H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
&= I(\hat{\mathbf{X}}; \mathbf{X}) \\
&= H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
&= \sum_{k=1}^{n} H(X_k) \\
&\quad - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
&\geq \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
&= \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
&= \sum_{k=1}^{n} \underline{I(X_k; \hat{X}_k)} \\
&\geq \sum_{k=1}^{n} \underline{R_I(Ed(X_k, \hat{X}_k))}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n}\log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
&\geq \log M \\
&\geq H(f(\mathbf{X})) \\
&\geq H(g(f(\mathbf{X}))) \\
&= H(\hat{\mathbf{X}}) \\
&= H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
&= I(\hat{\mathbf{X}}; \mathbf{X}) \\
&= H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
&= \sum_{k=1}^{n} H(X_k) \\
&\quad - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
&\geq \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
&= \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
&= \sum_{k=1}^{n} \underline{I(X_k; \hat{X}_k)} \\
&\geq \sum_{k=1}^{n} R_I(\underline{Ed(X_k, \hat{X}_k)})
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R + \epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \quad & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \quad & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= \quad & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq \quad & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= \quad & n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right]
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R + \epsilon) & \\
\geq \; & \log M \\
\geq \; & H(f(\mathbf{X})) \\
\geq \; & H(g(f(\mathbf{X}))) \\
= \; & H(\hat{\mathbf{X}}) \\
= \; & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \; & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \; & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \; & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \; & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \; & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= \; & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq \; & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= \; & n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
\geq \; & n R_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right)
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n}\log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$
\begin{aligned}
n(R + \epsilon)&\\
&\geq \log M\\
&\geq H(f(\mathbf{X}))\\
&\geq H(g(f(\mathbf{X})))\\
&= H(\hat{\mathbf{X}})\\
&= H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X})\\
&= I(\hat{\mathbf{X}}; \mathbf{X})\\
&= H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}})\\
&= \sum_{k=1}^{n} H(X_k)\\
&\quad - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1})\\
&\geq \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k)\\
&= \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)]\\
&= \sum_{k=1}^{n} I(X_k; \hat{X}_k)\\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right]\\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right)
\end{aligned}
$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \quad & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \quad & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= \quad & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq \quad & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= \quad & n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
\geq \quad & n R_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right)
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq \;& \log M \\
\geq \;& H(f(\mathbf{X})) \\
\geq \;& H(g(f(\mathbf{X}))) \\
= \;& H(\hat{\mathbf{X}}) \\
= \;& H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \;& I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \;& H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \;& \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \;& \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \;& \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= \;& \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq \;& \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= \;& n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
\geq \;& nR_I\left(\underline{\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)}\right)
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R + \epsilon) & \\
\geq\ & \log M \\
\geq\ & H(f(\mathbf{X})) \\
\geq\ & H(g(f(\mathbf{X}))) \\
=\ & H(\hat{\mathbf{X}}) \\
=\ & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
=\ & I(\hat{\mathbf{X}}; \mathbf{X}) \\
=\ & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
=\ & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq\ & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
=\ & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
=\ & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq\ & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
=\ & n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
\geq\ & nR_I \left(\underline{\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k)} \right) \\
=\ & nR_I(\underline{Ed(\mathbf{X}, \hat{\mathbf{X}})}).
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R + \epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \quad & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \quad & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= \quad & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq \quad & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= \quad & n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
\geq \quad & n R_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right) \\
= \quad & n R_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
&\geq \log M \\
&\geq H(f(\mathbf{X})) \\
&\geq H(g(f(\mathbf{X}))) \\
&= H(\hat{\mathbf{X}}) \\
&= H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
&= I(\hat{\mathbf{X}}; \mathbf{X}) \\
&= H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
&= \sum_{k=1}^{n} H(X_k) \\
&\quad - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
&\geq \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
&= \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
&= \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
&= n\left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n}\log M \le R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \le \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\ge\ & \log M \\
\ge\ & H(f(\mathbf{X})) \\
\ge\ & H(g(f(\mathbf{X}))) \\
=\ & H(\hat{\mathbf{X}}) \\
=\ & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
=\ & I(\hat{\mathbf{X}}; \mathbf{X}) \\
=\ & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
=\ & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\ge\ & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
=\ & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
=\ & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\ge\ & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
=\ & n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
\ge\ & nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
=\ & nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$Ed(\mathbf{X}, \hat{\mathbf{X}})$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \le R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \le \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
& n(R+\epsilon) \\
& \quad \ge \quad \log M \\
& \quad \ge \quad H(f(\mathbf{X})) \\
& \quad \ge \quad H(g(f(\mathbf{X}))) \\
& \quad = \quad H(\hat{\mathbf{X}}) \\
& \quad = \quad H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
& \quad = \quad I(\hat{\mathbf{X}}; \mathbf{X}) \\
& \quad = \quad H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
& \quad = \quad \sum_{k=1}^{n} H(X_k) \\
& \qquad\qquad - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
& \quad \ge \quad \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
& \quad = \quad \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
& \quad = \quad \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
& \quad \ge \quad \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
& \quad = \quad n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
& \quad \ge \quad nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
& \quad = \quad nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
& Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
& \quad = \quad E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
& \qquad\quad +E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \le D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \le D + \epsilon\}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R + \epsilon) & \\
\geq\ & \log M \\
\geq\ & H(f(\mathbf{X})) \\
\geq\ & H(g(f(\mathbf{X}))) \\
=\ & H(\hat{\mathbf{X}}) \\
=\ & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
=\ & I(\hat{\mathbf{X}}; \mathbf{X}) \\
=\ & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
=\ & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq\ & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
=\ & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
=\ & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq\ & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
=\ & n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
\geq\ & n R_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right) \\
=\ & n R_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
& Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
&= E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
&\quad + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \quad & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \quad & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= \quad & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq \quad & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= \quad & n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
\geq \quad & nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
= \quad & nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
& Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
= \quad & E[d(\mathbf{X}, \hat{\mathbf{X}})|\underline{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon}]\Pr\{\underline{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon}\} \\
& + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R + \epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \quad & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \quad & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= \quad & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq \quad & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= \quad & n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
\geq \quad & n R_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right) \\
= \quad & n R_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x, \hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
& Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
& \quad = E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
& \quad\quad + E[d(\mathbf{X}, \hat{\mathbf{X}})|\underline{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon}]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n}\log M \le R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \le \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) &\\
\ge\ & \log M \\
\ge\ & H(f(\mathbf{X})) \\
\ge\ & H(g(f(\mathbf{X}))) \\
=\ & H(\hat{\mathbf{X}}) \\
=\ & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
=\ & I(\hat{\mathbf{X}}; \mathbf{X}) \\
=\ & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
=\ & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\ge\ & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
=\ & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
=\ & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\ge\ & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
=\ & n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
\ge\ & nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
=\ & nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
& Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
&= E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
&\quad + E[d(\mathbf{X}, \hat{\mathbf{X}})|\underline{d(\mathbf{X}, \hat{\mathbf{X}}) \le D + \epsilon}]\Pr\{\underline{d(\mathbf{X}, \hat{\mathbf{X}}) \le D + \epsilon}\}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq \;& \log M \\
\geq \;& H(f(\mathbf{X})) \\
\geq \;& H(g(f(\mathbf{X}))) \\
= \;& H(\hat{\mathbf{X}}) \\
= \;& H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \;& I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \;& H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \;& \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \;& \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \;& \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= \;& \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq \;& \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= \;& n\left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
\geq \;& nR_I\left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
= \;& nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x, \hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
& Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
& = E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
& \quad + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \quad & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \quad & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= \quad & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq \quad & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= \quad & n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
\geq \quad & nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
= \quad & nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
Ed(\mathbf{X}, \hat{\mathbf{X}}) & \\
= \quad & E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
& + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\} \\
\leq \quad & d_{max} \cdot \epsilon + (D + \epsilon) \cdot 1
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq\ & \log M \\
\geq\ & H(f(\mathbf{X})) \\
\geq\ & H(g(f(\mathbf{X}))) \\
=\ & H(\hat{\mathbf{X}}) \\
=\ & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
=\ & I(\hat{\mathbf{X}}; \mathbf{X}) \\
=\ & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
=\ & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq\ & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
=\ & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
=\ & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq\ & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
=\ & n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
\geq\ & nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
=\ & nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
& Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
=\ & E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon]\underline{\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon\}} \\
& +E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon\} \\
\leq\ & d_{max} \cdot \epsilon + (D+\epsilon) \cdot 1
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n}\log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq & \log M \\
\geq & H(f(\mathbf{X})) \\
\geq & H(g(f(\mathbf{X}))) \\
= & H(\hat{\mathbf{X}}) \\
= & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= & n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
\geq & nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
= & nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
Ed(\mathbf{X}, \hat{\mathbf{X}}) & \\
= & E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon\} \\
& + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon\} \\
\leq & d_{max} \cdot \epsilon + (D+\epsilon) \cdot 1
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R + \epsilon) & \\
\geq\ & \log M \\
\geq\ & H(f(\mathbf{X})) \\
\geq\ & H(g(f(\mathbf{X}))) \\
=\ & H(\hat{\mathbf{X}}) \\
=\ & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
=\ & I(\hat{\mathbf{X}}; \mathbf{X}) \\
=\ & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
=\ & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq\ & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
=\ & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
=\ & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq\ & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
=\ & n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
\geq\ & n R_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right) \\
=\ & n R_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x, \hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
& Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
& = E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
& \quad + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\} \\
& \leq d_{max} \cdot \epsilon + (D + \epsilon) \cdot 1
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq\ & \log M \\
\geq\ & H(f(\mathbf{X})) \\
\geq\ & H(g(f(\mathbf{X}))) \\
=\ & H(\hat{\mathbf{X}}) \\
=\ & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
=\ & I(\hat{\mathbf{X}}; \mathbf{X}) \\
=\ & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
=\ & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq\ & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
=\ & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
=\ & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq\ & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
=\ & n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
\geq\ & nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
=\ & nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
Ed(\mathbf{X}, \hat{\mathbf{X}}) & \\
=\ & E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
& + \underline{E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]}\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\} \\
\leq\ & d_{max} \cdot \epsilon + (D + \epsilon) \cdot 1
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R + \epsilon) & \\
\geq \quad & \log M \\
\geq \quad & H(f(\mathbf{X})) \\
\geq \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq \quad & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \quad & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= \quad & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq \quad & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= \quad & n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
\geq \quad & n R_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right) \\
= \quad & n R_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x, \hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
Ed(\mathbf{X}, \hat{\mathbf{X}}) & \\
= \quad & E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
& + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\} \\
\leq \quad & d_{max} \cdot \epsilon + (D + \epsilon) \cdot 1
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\geq & \quad \log M \\
\geq & \quad H(f(\mathbf{X})) \\
\geq & \quad H(g(f(\mathbf{X}))) \\
= & \quad H(\hat{\mathbf{X}}) \\
= & \quad H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= & \quad I(\hat{\mathbf{X}}; \mathbf{X}) \\
= & \quad H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= & \quad \sum_{k=1}^{n} H(X_k) \\
& \quad - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\geq & \quad \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= & \quad \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= & \quad \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\geq & \quad \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= & \quad n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
\geq & \quad nR_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right) \\
= & \quad nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x, \hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
Ed(\mathbf{X}, \hat{\mathbf{X}}) & \\
= & \quad E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
& \quad + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]\underline{\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\}} \\
\leq & \quad d_{max} \cdot \epsilon + (D + \epsilon) \cdot 1
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
& n(R + \epsilon) \\
& \quad \geq \log M \\
& \quad \geq H(f(\mathbf{X})) \\
& \quad \geq H(g(f(\mathbf{X}))) \\
& \quad = H(\hat{\mathbf{X}}) \\
& \quad = H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
& \quad = I(\hat{\mathbf{X}}; \mathbf{X}) \\
& \quad = H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
& \quad = \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
& \quad \geq \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
& \quad = \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
& \quad = \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
& \quad \geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
& \quad = n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
& \quad \geq n R_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right) \\
& \quad = n R_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
& Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
& \quad = E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
& \quad\quad + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]\underline{\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\}} \\
& \quad \leq d_{max} \cdot \epsilon + (D + \epsilon) \cdot \underline{1}
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n}\log M \le R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \le \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\ge\ & \log M \\
\ge\ & H(f(\mathbf{X})) \\
\ge\ & H(g(f(\mathbf{X}))) \\
=\ & H(\hat{\mathbf{X}}) \\
=\ & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
=\ & I(\hat{\mathbf{X}}; \mathbf{X}) \\
=\ & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
=\ & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\ge\ & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
=\ & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
=\ & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\ge\ & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
=\ & n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
\ge\ & nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
=\ & nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
Ed(\mathbf{X}, \hat{\mathbf{X}}) & \\
=\ & E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon\} \\
& + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \le D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \le D+\epsilon\} \\
\le\ & d_{max}\cdot\epsilon + (D+\epsilon)\cdot 1 \\
=\ & D + (d_{max}+1)\epsilon.
\end{aligned}$$

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n}\log M \leq R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \leq \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
&n(R+\epsilon) \\
&\quad\geq \log M \\
&\quad\geq H(f(\mathbf{X})) \\
&\quad\geq H(g(f(\mathbf{X}))) \\
&\quad= H(\hat{\mathbf{X}}) \\
&\quad= H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
&\quad= I(\hat{\mathbf{X}}; \mathbf{X}) \\
&\quad= H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
&\quad= \sum_{k=1}^{n} H(X_k) \\
&\qquad - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
&\quad\geq \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
&\quad= \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
&\quad= \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
&\quad\geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
&\quad= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
&\quad\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
&\quad= nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
&Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
&\quad= E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
&\qquad + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\} \\
&\quad\leq d_{max} \cdot \epsilon + (D + \epsilon) \cdot 1 \\
&\quad= D + (d_{max} + 1)\epsilon.
\end{aligned}$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D + \epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

The Converse

1. Let (R, D) be any achievable rate-distortion pair, i.e., for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M \le R + \epsilon$$

and

$$\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \le \epsilon,$$

where $\hat{\mathbf{X}} = g(f(\mathbf{X}))$.

2. Then

$$\begin{aligned}
n(R+\epsilon) & \\
\ge \quad & \log M \\
\ge \quad & H(f(\mathbf{X})) \\
\ge \quad & H(g(f(\mathbf{X}))) \\
= \quad & H(\hat{\mathbf{X}}) \\
= \quad & H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}) \\
= \quad & I(\hat{\mathbf{X}}; \mathbf{X}) \\
= \quad & H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\
= \quad & \sum_{k=1}^{n} H(X_k) \\
& - \sum_{k=1}^{n} H(X_k|\hat{\mathbf{X}}, X_1, X_2, \cdots, X_{k-1}) \\
\ge \quad & \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \\
= \quad & \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
= \quad & \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
\ge \quad & \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
= \quad & n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
\ge \quad & n R_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right) \\
= \quad & n R_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}$$

3. Let $d_{max} = \max_{x, \hat{x}} d(x, \hat{x})$. Then

$$\begin{aligned}
& Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
= \quad & E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
& + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \le D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \le D + \epsilon\} \\
\le \quad & d_{max} \cdot \epsilon + (D + \epsilon) \cdot 1 \\
= \quad & D + (d_{max} + 1)\epsilon.
\end{aligned}$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D + \epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

The Converse

$$
\begin{aligned}
&= \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] \\
&= \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
&= n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
&\geq nR_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right) \\
&= nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
&\quad = E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
&\quad\quad + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\} \\
&\quad \leq d_{max} \cdot \epsilon + (D + \epsilon) \cdot 1 \\
&\quad = D + (d_{max} + 1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D + \epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
&\quad = E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon\} \\
&\qquad +E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\quad \leq d_{max} \cdot \epsilon + (D+\epsilon)\cdot 1 \\
&\quad = D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k;\hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k,\hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X},\hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x,\hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X},\hat{\mathbf{X}}) \\
&\quad= E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon\} \\
&\qquad +E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\quad\leq d_{max}\cdot\epsilon + (D+\epsilon)\cdot 1 \\
&\quad= D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
&= n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
&\geq nR_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right) \\
&= nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
&\quad = E[d(\mathbf{X}, \hat{\mathbf{X}}) | d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon] \Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon\} \\
&\qquad + E[d(\mathbf{X}, \hat{\mathbf{X}}) | d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon] \Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\quad \leq d_{max} \cdot \epsilon + (D+\epsilon) \cdot 1 \\
&\quad = D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
&= n \left[\frac{1}{n} \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \right] \\
&\geq n R_I \left(\frac{1}{n} \sum_{k=1}^{n} Ed(X_k, \hat{X}_k) \right) \\
&= n R_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
&\quad = E[d(\mathbf{X}, \hat{\mathbf{X}}) | d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon] \Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
&\qquad + E[d(\mathbf{X}, \hat{\mathbf{X}}) | d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon] \Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\} \\
&\quad \leq d_{max} \cdot \epsilon + (D + \epsilon) \cdot 1 \\
&\quad = D + (d_{max} + 1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D + \epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

The Converse

$$\begin{aligned} n(R+\epsilon) &\geq \cdots \\ &= \sum_{k=1}^{n} I(X_k;\hat{X}_k) \\ &\geq \sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k)) \\ &= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k))\right] \\ &\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k,\hat{X}_k)\right) \\ &= nR_I(Ed(\mathbf{X},\hat{\mathbf{X}})). \end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x,\hat{x})$. Then

$$\begin{aligned} &Ed(\mathbf{X},\hat{\mathbf{X}}) \\ &\quad= E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon\} \\ &\qquad + E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon\} \\ &\quad\leq d_{max}\cdot\epsilon + (D+\epsilon)\cdot 1 \\ &\quad= D + (d_{max}+1)\epsilon. \end{aligned}$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$R+\epsilon \geq R_I(Ed(\mathbf{X},\hat{\mathbf{X}}))$$

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$
\begin{aligned}
& Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
&\quad = E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon\} \\
&\qquad + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\quad \leq d_{max} \cdot \epsilon + (D+\epsilon) \cdot 1 \\
&\quad = D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$
R+\epsilon \geq R_I(Ed(\mathbf{X}, \hat{\mathbf{X}}))
$$

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k;\hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k,\hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X},\hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x,\hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X},\hat{\mathbf{X}}) \\
&= E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon\} \\
&\quad + E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\leq d_{max}\cdot\epsilon + (D+\epsilon)\cdot 1 \\
&= D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$
R+\epsilon \geq R_I(Ed(\mathbf{X},\hat{\mathbf{X}}))
$$

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k;\hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k,\hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X},\hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x,\hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X},\hat{\mathbf{X}}) \\
&= E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon\} \\
&\quad + E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\leq d_{max}\cdot\epsilon + (D+\epsilon)\cdot 1 \\
&= D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$
\begin{aligned}
R+\epsilon &\geq R_I(Ed(\mathbf{X},\hat{\mathbf{X}})) \\
&\geq R_I(D+(d_{max}+1)\epsilon),
\end{aligned}
$$

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
&= E[d(\mathbf{X}, \hat{\mathbf{X}}) | d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D + \epsilon\} \\
&\quad + E[d(\mathbf{X}, \hat{\mathbf{X}}) | d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D + \epsilon\} \\
&\leq d_{max} \cdot \epsilon + (D + \epsilon) \cdot 1 \\
&= D + (d_{max} + 1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D + \epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$
\begin{aligned}
R + \epsilon &\geq R_I(Ed(\mathbf{X}, \hat{\mathbf{X}})) \\
&\geq R_I(D + (d_{max} + 1)\epsilon),
\end{aligned}
$$

because $R_I(D)$ is non-increasing in D.

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
&\quad = E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon\} \\
&\qquad + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\quad \leq d_{max} \cdot \epsilon + (D+\epsilon) \cdot 1 \\
&\quad = D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$
\begin{aligned}
R+\epsilon &\geq R_I(Ed(\mathbf{X}, \hat{\mathbf{X}})) \\
&\geq R_I(D + (d_{max}+1)\epsilon),
\end{aligned}
$$

because $R_I(D)$ is non-increasing in D.

The Converse

$$\begin{aligned} n(R+\epsilon) &\geq \cdots \\ &= \sum_{k=1}^{n} I(X_k;\hat{X}_k) \\ &\geq \sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k)) \\ &= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k))\right] \\ &\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k,\hat{X}_k)\right) \\ &= nR_I(Ed(\mathbf{X},\hat{\mathbf{X}})). \end{aligned}$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x,\hat{x})$. Then

$$\begin{aligned} &Ed(\mathbf{X},\hat{\mathbf{X}}) \\ &= E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon\} \\ &\quad +E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon\} \\ &\leq d_{max}\cdot\epsilon + (D+\epsilon)\cdot 1 \\ &= D + (d_{max}+1)\epsilon. \end{aligned}$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$\begin{aligned} R+\epsilon &\geq R_I(Ed(\mathbf{X},\hat{\mathbf{X}})) \\ &\geq R_I(D+(d_{max}+1)\epsilon), \end{aligned}$$

because $R_I(D)$ is non-increasing in D.

5. $R_I(D)$ convex implies it is continuous in D. Letting $\epsilon \to 0$, we have

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k;\hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k,\hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X},\hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x,\hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X},\hat{\mathbf{X}}) \\
&\quad = E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon\} \\
&\qquad +E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\quad \leq d_{max}\cdot\epsilon + (D+\epsilon)\cdot 1 \\
&\quad = D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$
\begin{aligned}
R+\epsilon &\geq R_I(Ed(\mathbf{X},\hat{\mathbf{X}})) \\
&\geq R_I(D+(d_{max}+1)\epsilon),
\end{aligned}
$$

because $R_I(D)$ is non-increasing in D.

5. $R_I(D)$ convex implies it is continuous in D. Letting $\epsilon \to 0$, we have

$$
R \geq \lim_{\epsilon\to 0} R_I(D+(d_{max}+1)\epsilon)
$$

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k;\hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k,\hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X},\hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x,\hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X},\hat{\mathbf{X}}) \\
&\quad= E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon\} \\
&\qquad +E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\quad\leq d_{max}\cdot\epsilon + (D+\epsilon)\cdot 1 \\
&\quad= D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$
\begin{aligned}
R+\epsilon &\geq R_I(Ed(\mathbf{X},\hat{\mathbf{X}})) \\
&\geq R_I(D+(d_{max}+1)\epsilon),
\end{aligned}
$$

because $R_I(D)$ is non-increasing in D.

5. $R_I(D)$ convex implies it is continuous in D. Letting $\epsilon \to 0$, we have

$$
\begin{aligned}
R &\geq \lim_{\epsilon\to 0} R_I(D+(d_{max}+1)\epsilon) \\
&= R_I\left(D+(d_{max}+1)\lim_{\epsilon\to 0}\epsilon\right)
\end{aligned}
$$

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k; \hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k, \hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k, \hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X}, \hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x, \hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X}, \hat{\mathbf{X}}) \\
&\quad = E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) > D+\epsilon\} \\
&\qquad + E[d(\mathbf{X}, \hat{\mathbf{X}})|d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X}, \hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\quad \leq d_{max} \cdot \epsilon + (D+\epsilon) \cdot 1 \\
&\quad = D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$
\begin{aligned}
R+\epsilon &\geq R_I(Ed(\mathbf{X}, \hat{\mathbf{X}})) \\
&\geq R_I(D + (d_{max}+1)\epsilon),
\end{aligned}
$$

because $R_I(D)$ is non-increasing in D.

5. $R_I(D)$ convex implies it is continuous in D. Letting $\epsilon \to 0$, we have

$$
\begin{aligned}
R &\geq \lim_{\epsilon\to 0} R_I(D + (d_{max}+1)\epsilon) \\
&= R_I\left(D + (d_{max}+1)\lim_{\epsilon\to 0}\epsilon\right)
\end{aligned}
$$

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k;\hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k,\hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X},\hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x,\hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X},\hat{\mathbf{X}}) \\
&\quad = E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon\} \\
&\qquad + E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\quad \leq d_{max}\cdot\epsilon + (D+\epsilon)\cdot 1 \\
&\quad = D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$
\begin{aligned}
R+\epsilon &\geq R_I(Ed(\mathbf{X},\hat{\mathbf{X}})) \\
&\geq R_I(D+(d_{max}+1)\epsilon),
\end{aligned}
$$

because $R_I(D)$ is non-increasing in D.

5. $R_I(D)$ convex implies it is continuous in D. Letting $\epsilon \to 0$, we have

$$
\begin{aligned}
R &\geq \lim_{\epsilon\to 0} R_I(D+(d_{max}+1)\epsilon) \\
&= R_I\left(D+(d_{max}+1)\lim_{\epsilon\to 0}\epsilon\right) \\
&= R_I(D).
\end{aligned}
$$

The Converse

$$
\begin{aligned}
n(R+\epsilon) &\geq \cdots \\
&= \sum_{k=1}^{n} I(X_k;\hat{X}_k) \\
&\geq \sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k)) \\
&= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I(Ed(X_k,\hat{X}_k))\right] \\
&\geq nR_I\left(\frac{1}{n}\sum_{k=1}^{n} Ed(X_k,\hat{X}_k)\right) \\
&= nR_I(Ed(\mathbf{X},\hat{\mathbf{X}})).
\end{aligned}
$$

3. Let $d_{max} = \max_{x,\hat{x}} d(x,\hat{x})$. Then

$$
\begin{aligned}
&Ed(\mathbf{X},\hat{\mathbf{X}}) \\
&\quad= E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) > D+\epsilon\} \\
&\qquad +E[d(\mathbf{X},\hat{\mathbf{X}})|d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon]\Pr\{d(\mathbf{X},\hat{\mathbf{X}}) \leq D+\epsilon\} \\
&\quad\leq d_{max}\cdot\epsilon + (D+\epsilon)\cdot 1 \\
&\quad= D + (d_{max}+1)\epsilon.
\end{aligned}
$$

That is, if the probability that the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ exceeds $D+\epsilon$ is small, then the expected average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ can exceed D only by a small amount.

4. Therefore,

$$
\begin{aligned}
R+\epsilon &\geq R_I(Ed(\mathbf{X},\hat{\mathbf{X}})) \\
&\geq R_I(D+(d_{max}+1)\epsilon),
\end{aligned}
$$

because $R_I(D)$ is non-increasing in D.

5. $R_I(D)$ convex implies it is continuous in D. Letting $\epsilon \to 0$, we have

$$
\begin{aligned}
R &\geq \lim_{\epsilon\to 0} R_I(D+(d_{max}+1)\epsilon) \\
&= R_I\left(D + (d_{max}+1)\lim_{\epsilon\to 0}\epsilon\right) \\
&= R_I(D).
\end{aligned}
$$

Finally, minimize R over all achievable pairs (R, D) for a fixed D to obtain $R(D) \geq R_I(D)$.