

Notions of Entropy for Information Channels

Input/Output Entropy: Average information into/out-of Γ .

$$H(\mathcal{A}) = \sum_i p_i \log \frac{1}{p_i} \quad , \quad H(\mathcal{B}) = \sum_j q_j \log \frac{1}{q_j}$$

Lecture 6:
Information
Channel Measures:
 H, I, and C

System Entropies
Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
 Definition
 For BSCs
 Channel Capacity

Epilogue

Notions of Entropy for Information Channels

Input/Output Entropy: Average information into/out-of Γ .

$$H(\mathcal{A}) = \sum_i p_i \log \frac{1}{p_i} \quad , \quad H(\mathcal{B}) = \sum_j q_j \log \frac{1}{q_j}$$

Conditional Entropy: Uncertainty of \mathcal{A} , when b_j known, and vice versa.

Lecture 6:
Information
Channel Measures:
 H, I, and C

System Entropies
Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
 Definition
 For BSCs
 Channel Capacity

Epilogue

Notions of Entropy for Information Channels

Input/Output Entropy: Average information into/out-of Γ .

$$H(\mathcal{A}) = \sum_i p_i \log \frac{1}{p_i} \quad , \quad H(\mathcal{B}) = \sum_j q_j \log \frac{1}{q_j}$$

Conditional Entropy: Uncertainty of \mathcal{A} , when b_j known, and vice versa.

$$H(\mathcal{A} | b_j) = \sum_i \Pr(a_i | b_j) \log \frac{1}{\Pr(a_i | b_j)} = \sum_i Q_{ij} \log \frac{1}{Q_{ij}}$$

$$H(\mathcal{B} | a_i) = \sum_j P_{ij} \log \frac{1}{P_{ij}} \quad , \quad \text{in a similar manner.}$$

Lecture 6:
Information
Channel Measures:
H, I, and C

System Entropies
Definitions
For BSCs

Shannon's First
Theorem for Γ
The Idea
Problems

Mutual Information
Definition
For BSCs
Channel Capacity

Epilogue

Notions of Entropy for Information Channels

Input/Output Entropy: Average information into/out-of Γ .

$$H(\mathcal{A}) = \sum_i p_i \log \frac{1}{p_i} \quad , \quad H(\mathcal{B}) = \sum_j q_j \log \frac{1}{q_j}$$

Conditional Entropy: Uncertainty of \mathcal{A} , when b_j known, and vice versa.

$$H(\mathcal{A} | b_j) = \sum_i \Pr(a_i | b_j) \log \frac{1}{\Pr(a_i | b_j)} = \sum_i Q_{ij} \log \frac{1}{Q_{ij}}$$

$$H(\mathcal{B} | a_i) = \sum_j P_{ij} \log \frac{1}{P_{ij}} \quad , \quad \text{in a similar manner.}$$

Equivocation: Average Uncertainty of \mathcal{A} w.r.t. \mathcal{B} , and vice versa.

Lecture 6:
Information
Channel Measures:
H, I, and C

System Entropies
Definitions
For BSCs

Shannon's First
Theorem for Γ
The Idea
Problems

Mutual Information
Definition
For BSCs
Channel Capacity

Epilogue

Notions of Entropy for Information Channels

Input/Output Entropy: Average information into/out-of Γ .

$$H(\mathcal{A}) = \sum_i p_i \log \frac{1}{p_i} \quad , \quad H(\mathcal{B}) = \sum_j q_j \log \frac{1}{q_j}$$

Conditional Entropy: Uncertainty of \mathcal{A} , when b_j known, and vice versa.

$$H(\mathcal{A} | b_j) = \sum_i \Pr(a_i | b_j) \log \frac{1}{\Pr(a_i | b_j)} = \sum_i Q_{ij} \log \frac{1}{Q_{ij}}$$

$$H(\mathcal{B} | a_i) = \sum_j P_{ij} \log \frac{1}{P_{ij}} \quad , \text{ in a similar manner.}$$

Equivocation: Average Uncertainty of \mathcal{A} w.r.t. \mathcal{B} , and vice versa.

$$\begin{aligned} H(\mathcal{A} | \mathcal{B}) &= \overbrace{\sum_j q_j} H(\mathcal{A} | b_j) = \sum_j q_j \left(\sum_i Q_{ij} \log \frac{1}{Q_{ij}} \right) \\ &= \sum_i \sum_j R_{ij} \log \frac{1}{Q_{ij}} \quad , \text{ since } q_j Q_{ij} = R_{ij}. \end{aligned}$$

$$H(\mathcal{B} | \mathcal{A}) = \sum_i \sum_j R_{ij} \log \frac{1}{P_{ij}} \quad , \text{ in a similar manner.}$$

Joint Entropy: Average uncertainty when observer guesses i/o of Γ .

Lecture 6:
Information
Channel Measures:
H, I, and C

System Entropies

Definitions

For BSCs

Shannon's First
Theorem for Γ

The Idea
Problems

Mutual Information
Definition
For BSCs
Channel Capacity

Epilogue

Joint Entropy: Average uncertainty when observer guesses i/o of Γ .

$$H(\mathcal{A}, \mathcal{B}) = \sum_{i,j} \Pr(a_i, b_j) \log \frac{1}{\Pr(a_i, b_j)} = \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}}$$

Lecture 6:
Information
Channel Measures:
 H, I, and C

System Entropies

Definitions

For BSCs

Shannon's First

Theorem for Γ

The Idea

Problems

Mutual Information

Definition

For BSCs

Channel Capacity

Epilogue

Joint Entropy: Average uncertainty when observer guesses i/o of Γ .

$$\begin{aligned} H(\mathcal{A}, \mathcal{B}) &= \sum_{i,j} \Pr(a_i, b_j) \log \frac{1}{\Pr(a_i, b_j)} = \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}} \\ &= \sum_i \sum_j R_{ij} \log \frac{1}{p_i} + \sum_i \sum_j R_{ij} \log \frac{1}{p_{ij}} \text{ , as } R_{ij} = p_i p_{ij}. \end{aligned}$$

Lecture 6:
Information
Channel Measures:
 H, I, and C

System Entropies
Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
 Definition
 For BSCs
 Channel Capacity

Epilogue

Joint Entropy: Average uncertainty when observer guesses i/o of Γ .

$$\begin{aligned}
 H(\mathcal{A}, \mathcal{B}) &= \sum_{i,j} \Pr(a_i, b_j) \log \frac{1}{\Pr(a_i, b_j)} = \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}} \\
 &= \sum_i \sum_j R_{ij} \log \frac{1}{p_i} + \sum_i \sum_j R_{ij} \log \frac{1}{p_{ij}} , \text{ as } R_{ij} = p_i p_{ij}. \\
 &= H(\mathcal{A}) + H(\mathcal{A} \mid \mathcal{B}) , \text{ from prior definitions.}
 \end{aligned}$$

Lecture 6:
Information
Channel Measures:
 H, I, and C

System Entropies

Definitions

For BSCs

Shannon's First
 Theorem for Γ

The Idea
 Problems

Mutual Information

Definition

For BSCs

Channel Capacity

Epilogue

Joint Entropy: Average uncertainty when observer guesses i/o of Γ .

$$\begin{aligned}
 H(\mathcal{A}, \mathcal{B}) &= \sum_{i,j} \Pr(a_i, b_j) \log \frac{1}{\Pr(a_i, b_j)} = \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}} \\
 &= \sum_i \sum_j R_{ij} \log \frac{1}{p_i} + \sum_i \sum_j R_{ij} \log \frac{1}{p_{ij}} , \text{ as } R_{ij} = p_i p_{ij}. \\
 &= H(\mathcal{A}) + H(\mathcal{A} \mid \mathcal{B}) , \text{ from prior definitions.} \\
 &= H(\mathcal{B}) + H(\mathcal{B} \mid \mathcal{A}) , \text{ as } R_{ij} = q_j Q_{ij}
 \end{aligned}$$

$H(\mathcal{B} \mid \mathcal{A})$ is thus the *extra* information conveyed by \mathcal{B} when \mathcal{A} is known (if Γ reliable, then 0). Like $\Pr(\mathcal{A} \cup \mathcal{B}) = \Pr(\mathcal{A}) + \Pr(\mathcal{B} \setminus \mathcal{A})$. And $|\mathcal{A} \cup \mathcal{B}| = |\mathcal{A}| + |\mathcal{B} \setminus \mathcal{A}|$. And vice versa.

Joint Entropy: Average uncertainty when observer guesses i/o of Γ .

$$\begin{aligned}
 H(\mathcal{A}, \mathcal{B}) &= \sum_{i,j} \Pr(a_i, b_j) \log \frac{1}{\Pr(a_i, b_j)} = \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}} \\
 &= \sum_i \sum_j R_{ij} \log \frac{1}{p_i} + \sum_i \sum_j R_{ij} \log \frac{1}{p_{ij}}, \text{ as } R_{ij} = p_i p_{ij}. \\
 &= H(\mathcal{A}) + H(\mathcal{A} \mid \mathcal{B}), \text{ from prior definitions.} \\
 &= H(\mathcal{B}) + H(\mathcal{B} \mid \mathcal{A}), \text{ as } R_{ij} = q_j Q_{ij}
 \end{aligned}$$

$H(\mathcal{B} \mid \mathcal{A})$ is thus the *extra* information conveyed by \mathcal{B} when \mathcal{A} is known (if Γ reliable, then 0). Like $\Pr(A \cup B) = \Pr(A) + \Pr(B \setminus A)$. And $|A \cup B| = |A| + |B \setminus A|$. And vice versa.

Joint Entropy, Independent: If \mathcal{A}, \mathcal{B} are *statistically independent* ($R_{ij} = p_i q_j$),

$$\begin{aligned}
 H(\mathcal{A}, \mathcal{B}) &= \sum_{i,j} p_i q_j \log \frac{1}{p_i q_j} = \sum_i p_i \log \frac{1}{p_i} + \sum_j q_j \log \frac{1}{q_j} \\
 &= H(\mathcal{A}) + H(\mathcal{B})
 \end{aligned}$$

Like $|A \cup B| = |A| + |B|$ when A, B are disjoint.

Convexity of $H(p)$, and its implications

Recall that $(p_0, p_1) = (p, \bar{p})$ and $(q_0, q_1) = (q, \bar{q})$, and

$$P_{ij} = \begin{pmatrix} p & \bar{p} \\ \bar{p} & p \end{pmatrix}. \quad \text{so, } (q, \bar{q}) = (p, \bar{p}) \begin{pmatrix} p & \bar{p} \\ \bar{p} & p \end{pmatrix} = (pP + \bar{p}\bar{P}, p\bar{P} + \bar{p}P).$$

Lecture 6:
Information
Channel Measures:
H, I, and C

System Entropies
 Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
 Definition
 For BSCs
 Channel Capacity

Epilogue

Convexity of $H(p)$, and its implications

Recall that $(p_0, p_1) = (p, \bar{p})$ and $(q_0, q_1) = (q, \bar{q})$, and

$$P_{ij} = \begin{pmatrix} p & \bar{p} \\ \bar{p} & p \end{pmatrix}. \quad \text{so, } (q, \bar{q}) = (p, \bar{p}) \begin{pmatrix} p & \bar{p} \\ \bar{p} & p \end{pmatrix} = (pP + \bar{p}\bar{P}, p\bar{P} + \bar{p}P).$$

$$H(\mathcal{A}) = -p \log p - \bar{p} \log \bar{p} = H(p)$$

$$H(\mathcal{B}) = -q \log q - \bar{q} \log \bar{q} = H(q)$$

Lecture 6:
Information
Channel Measures:
 H , I , and C

System Entropies
 Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
 Definition
 For BSCs
 Channel Capacity

Epilogue

Convexity of $H(p)$, and its implications

Recall that $(p_0, p_1) = (p, \bar{p})$ and $(q_0, q_1) = (q, \bar{q})$, and

$$P_{ij} = \begin{pmatrix} p & \bar{p} \\ \bar{p} & p \end{pmatrix}. \quad \text{so, } (q, \bar{q}) = (p, \bar{p}) \begin{pmatrix} p & \bar{p} \\ \bar{p} & p \end{pmatrix} = (p^2 + \bar{p}^2, p\bar{p} + \bar{p}p).$$

$$H(\mathcal{A}) = -p \log p - \bar{p} \log \bar{p} = H(p)$$

$$H(\mathcal{B}) = -q \log q - \bar{q} \log \bar{q} = H(q)$$

Definition (Strictly Convex Function)

$f : [0, 1] \rightarrow \mathbb{R}$ is *strictly convex* if, when $a, b \in [0, 1]$ and $x = \lambda a + \bar{\lambda} b$ with $0 \leq \lambda \leq 1$, then

$$f(x) \geq \lambda f(a) + \bar{\lambda} f(b).$$

Convexity of $H(p)$, and its implications

Recall that $(p_0, p_1) = (p, \bar{p})$ and $(q_0, q_1) = (q, \bar{q})$, and

$$P_{ij} = \begin{pmatrix} p & \bar{p} \\ \bar{p} & p \end{pmatrix}. \quad \text{so, } (q, \bar{q}) = (p, \bar{p}) \begin{pmatrix} p & \bar{p} \\ \bar{p} & p \end{pmatrix} = (p^2 + \bar{p}^2, p\bar{p} + \bar{p}p).$$

$$H(\mathcal{A}) = -p \log p - \bar{p} \log \bar{p} = H(p)$$

$$H(\mathcal{B}) = -q \log q - \bar{q} \log \bar{q} = H(q)$$

Definition (Strictly Convex Function)

$f : [0, 1] \rightarrow \mathbb{R}$ is *strictly convex* if, when $a, b \in [0, 1]$ and $x = \lambda a + \bar{\lambda} b$ with $0 \leq \lambda \leq 1$, then

$$f(x) \geq \lambda f(a) + \bar{\lambda} f(b).$$

Corollary (Convexity of $H(p)$ (proof omitted; draw))

$H(p)$ is *strictly convex*.

Convexity of $H(p)$, and its implications

Recall that $(p_0, p_1) = (p, \bar{p})$ and $(q_0, q_1) = (q, \bar{q})$, and

$$p_{ij} = \begin{pmatrix} p & \bar{p} \\ \bar{p} & p \end{pmatrix}. \quad \text{so, } (q, \bar{q}) = (p, \bar{p}) \begin{pmatrix} p & \bar{p} \\ \bar{p} & p \end{pmatrix} = (pP + \bar{p}\bar{P}, p\bar{P} + \bar{p}P).$$

$$H(\mathcal{A}) = -p \log p - \bar{p} \log \bar{p} = H(p)$$

$$H(\mathcal{B}) = -q \log q - \bar{q} \log \bar{q} = H(q)$$

Definition (Strictly Convex Function)

$f : [0, 1] \rightarrow \mathbb{R}$ is *strictly convex* if, when $a, b \in [0, 1]$ and $x = \lambda a + \bar{\lambda} b$ with $0 \leq \lambda \leq 1$, then

$$f(x) \geq \lambda f(a) + \bar{\lambda} f(b).$$

Corollary (Convexity of $H(p)$ (proof omitted; draw))

$H(p)$ is *strictly convex*.

Set $a = p$, $b = \bar{p}$, $\lambda = P$. Then $x = pP + \bar{p}\bar{P} = q$. Thus

$$H(q) \geq H(p) \quad , \quad \text{that is,} \quad H(\mathcal{B}) \geq H(\mathcal{A}).$$

“=” only when $p = \frac{1}{2}$ or $P \in \{0, 1\}$. Sending through Γ increases uncertainty.

$$H(\mathcal{B} | \mathcal{A}) = \sum_{i,j} p_i P_{ij} \log \frac{1}{P_{ij}}$$

Lecture 6: Information Channel Measures: H, I, and C

System Entropies

Definitions

For BSCs

Shannon's First Theorem for Γ

The Idea Problems

Mutual Information

Definition

For BSCs

Channel Capacity

Epilogue

$$\begin{aligned}
H(\mathcal{B} \mid \mathcal{A}) &= \sum_{i,j} p_i P_{ij} \log \frac{1}{P_{ij}} \\
&= -pP \log P - p\bar{P} \log \bar{P} - \bar{p}\bar{P} \log \bar{P} - \bar{p}P \log P \\
&= -(p + \bar{p})\bar{P} \log \bar{P} - (p + \bar{p})P \log P \\
&= H(P)
\end{aligned}$$

So sender's uncertainty is the uncertainty in whether the symbol was sent correctly (minimal when $P \in \{0, 1\}$, maximal when $P = \frac{1}{2}$).

Lecture 6: Information Channel Measures: H, I, and C

System Entropies Definitions For BSCs

Shannon's First Theorem for Γ The Idea Problems

Mutual Information Definition For BSCs Channel Capacity

Epilogue

$$\begin{aligned}
H(\mathcal{B} \mid \mathcal{A}) &= \sum_{i,j} p_i P_{ij} \log \frac{1}{P_{ij}} \\
&= -pP \log P - p\bar{P} \log \bar{P} - \bar{p}\bar{P} \log \bar{P} - \bar{p}P \log P \\
&= -(p + \bar{p})\bar{P} \log \bar{P} - (p + \bar{p})P \log P \\
&= H(P)
\end{aligned}$$

So sender's uncertainty is the uncertainty in whether the symbol was sent correctly (minimal when $P \in \{0, 1\}$, maximal when $P = \frac{1}{2}$).

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B} \mid \mathcal{A}) = H(p) + H(P)$$

Lecture 6: Information Channel Measures: H, I, and C

System Entropies Definitions For BSCs

Shannon's First Theorem for Γ The Idea Problems

Mutual Information Definition For BSCs Channel Capacity

Epilogue

$$\begin{aligned}
H(\mathcal{B} \mid \mathcal{A}) &= \sum_{i,j} p_i P_{ij} \log \frac{1}{P_{ij}} \\
&= -pP \log P - p\bar{P} \log \bar{P} - \bar{p}\bar{P} \log \bar{P} - \bar{p}P \log P \\
&= -(p + \bar{p})\bar{P} \log \bar{P} - (p + \bar{p})P \log P \\
&= H(P)
\end{aligned}$$

So sender's uncertainty is the uncertainty in whether the symbol was sent correctly (minimal when $P \in \{0, 1\}$, maximal when $P = \frac{1}{2}$).

$$\begin{aligned}
H(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B} \mid \mathcal{A}) = H(p) + H(P) \\
&= H(\mathcal{B}) + H(\mathcal{A} \mid \mathcal{B}) = H(q) + H(\mathcal{A} \mid \mathcal{B})
\end{aligned}$$

Lecture 6: Information Channel Measures: H, I, and C

System Entropies Definitions For BSCs

Shannon's First Theorem for Γ The Idea Problems

Mutual Information Definition For BSCs Channel Capacity

Epilogue

$$\begin{aligned}
H(\mathcal{B} \mid \mathcal{A}) &= \sum_{i,j} p_i P_{ij} \log \frac{1}{P_{ij}} \\
&= -pP \log P - p\bar{P} \log \bar{P} - \bar{p}\bar{P} \log \bar{P} - \bar{p}P \log P \\
&= -(p + \bar{p})\bar{P} \log \bar{P} - (p + \bar{p})P \log P \\
&= H(P)
\end{aligned}$$

So sender's uncertainty is the uncertainty in whether the symbol was sent correctly (minimal when $P \in \{0, 1\}$, maximal when $P = \frac{1}{2}$).

$$\begin{aligned}
H(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B} \mid \mathcal{A}) = H(p) + H(P) \\
&= H(\mathcal{B}) + H(\mathcal{A} \mid \mathcal{B}) = H(q) + H(\mathcal{A} \mid \mathcal{B}) \\
H(\mathcal{A} \mid \mathcal{B}) &= H(p) + H(P) - H(q)
\end{aligned}$$

$$\begin{aligned}
 H(\mathcal{B} \mid \mathcal{A}) &= \sum_{i,j} p_i P_{ij} \log \frac{1}{P_{ij}} \\
 &= -pP \log P - p\bar{P} \log \bar{P} - \bar{p}\bar{P} \log \bar{P} - \bar{p}P \log P \\
 &= -(p + \bar{p})\bar{P} \log \bar{P} - (p + \bar{p})P \log P \\
 &= H(P)
 \end{aligned}$$

So sender's uncertainty is the uncertainty in whether the symbol was sent correctly (minimal when $P \in \{0, 1\}$, maximal when $P = \frac{1}{2}$).

$$\begin{aligned}
 H(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B} \mid \mathcal{A}) = H(p) + H(P) \\
 &= H(\mathcal{B}) + H(\mathcal{A} \mid \mathcal{B}) = H(q) + H(\mathcal{A} \mid \mathcal{B}) \\
 H(\mathcal{A} \mid \mathcal{B}) &= H(p) + H(P) - H(q)
 \end{aligned}$$

Since $q = pP + \bar{p}\bar{P}$; $\bar{P} \leq q \leq P$, convexity gives

$$H(P) \leq H(q) = H(\mathcal{B})$$

Thus

$$H(\mathcal{B} \mid \mathcal{A}) \leq H(\mathcal{B}) \quad , \text{ and from this, } H(\mathcal{A} \mid \mathcal{B}) \leq H(\mathcal{A})$$

“=” only when $P = \frac{1}{2}$ or $p \in \{0, 1\}$ in last two inequalities.

System Entropies
 Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
 Definition
 For BSCs
 Channel Capacity

Epilogue

Recap

Remember Shannon's First Theorem for \mathcal{S} :

- 1 $H(\mathcal{S}) \leq L(\mathcal{C})$ for any uniquely decodable encoding \mathcal{C} of \mathcal{S} .
- 2 We can get $L(\mathcal{C})$ arbitrarily close to $H(\mathcal{S})$ by extending \mathcal{S} .

In our new setting (note: $H(\mathcal{A} | \mathcal{B}) \leq H(\mathcal{A})$, proven later),

- 1 $H(\mathcal{A} | \mathcal{B}) \leq L(\mathcal{C})$ for any uniquely decodable encoding \mathcal{C} of \mathcal{A} .
- 2 We can get $L(\mathcal{C})$ arbitrarily close to $H(\mathcal{A} | \mathcal{B})$ by extending \mathcal{A} .

They mean two different things.

For \mathcal{S} : Information (code symbols) needed to know $s \in \mathcal{S}$.

Small $r \implies$ more information required.

For Γ : How much information (*redundancy*) we need to be *sure* which a_i my currently received b_j represents¹.

Γ unreliable \implies more information required.

Lecture 6:
Information
Channel Measures:
H, I, and C

System Entropies
Definitions
For BSCs

Shannon's First
Theorem for Γ
The Idea
Problems

Mutual Information
Definition
For BSCs
Channel Capacity

Epilogue

¹This extra information is not sent through Γ , but through a reliable channel.

The idea

- ➊ I send a symbol a_i . Receiver receives b_j .
- ➋ *I see what the receiver received b_j . Maybe the output symbol does not match a_i .*
- ➌ If not, I need to give the receiver more information *through a reliable medium*; enough information so that the receiver knows what word I sent originally.
- ➍ I do this by constructing a collection of Shannon-Fano codes; one \mathcal{C}_j for each b_j .
- ➎ When I sent a_i and receiver received b_j , I send the extra info as the code word $w_i \in \mathcal{C}_j$ ²

Theorem (Shannon's First Theorem for Γ)

If the output \mathcal{B} of Γ is known, then by encoding \mathcal{A}^n with n sufficiently large, one can find uniquely decodable encodings of the input \mathcal{A} with average word-lengths arbitrarily close to the equivocation $H(\mathcal{A} | \mathcal{B})$.

Proof similar as proof for \mathcal{S} , except we use $\Pr(a_i | b_j)$ instead of $\Pr(a_i)$.

²The word could, for instance, say "digit 2 and 6 got flipped; you should flip them back".

The Story Does Not End Here

This approach lets us communicate **compressed, error-free** data. And it tells us that we can do so up to the theoretical limit.

However, we assumed that

- we could see what the receiver received, so we could correct errors.
- we had a *reliable medium* to send the error-correction information to the receiver.

These assumptions make this result **impractical**.

Lecture 6:
Information
Channel Measures:
H, I, and C

System Entropies
Definitions
For BSCs

Shannon's First
Theorem for Γ
The Idea
Problems

Mutual Information
Definition
For BSCs
Channel Capacity

Epilogue

³Or decide, perhaps through empiric testing.

The Story Does Not End Here

This approach lets us communicate **compressed, error-free** data. And it tells us that we can do so up to the theoretical limit.

However, we assumed that

- we could see what the receiver received, so we could correct errors.
- we had a *reliable medium* to send the error-correction information to the receiver.

These assumptions make this result **impractical**. **Spoiler:** What we do instead is that we weave this error-correction data *into* the messages we send, so to say.

If we know³ the probability of our channel producing an error, we can weave sufficient error-correction into messages for the receiver to always decode correctly, provided the channel does not introduce more than the assumed nr. of errors.

The art is then to weave as *little* extra information into a message as possible, to achieve as *high* an error-correction as possible. **Shannon's Fundamental Theorem** gives theoretical minimum. We now start looking at preliminaries for its proof.

³Or decide, perhaps through empiric testing.

Definition

Recall that

- $H(\mathcal{A})$ is the uncertainty on \mathcal{A} when \mathcal{B} is *unknown*.
- $H(\mathcal{A} \mid \mathcal{B})$ is the uncertainty on \mathcal{A} when \mathcal{B} is *known*.

Mutual Information: Amount of uncertainty on \mathcal{A} *resolved* by knowing \mathcal{B} .

$$I(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) - H(\mathcal{A} \mid \mathcal{B})$$

Represents how much information \mathcal{A} and \mathcal{B} have in common.

Comparable to $|\mathcal{A} \cap \mathcal{B}| = |\mathcal{A}| - |\mathcal{A} \setminus \mathcal{B}|$. From

$$\begin{aligned} H(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B} \mid \mathcal{A}) \\ &= H(\mathcal{B}) + H(\mathcal{A} \mid \mathcal{B}), \end{aligned}$$

we get $I(\mathcal{A}, \mathcal{B}) = I(\mathcal{B}, \mathcal{A})$.

Interpretations

How much does reading the book tell you about the film? How much does watching the film tell you about the book? How much do notes tell you about the lecture? How much does the lecture tell about the notes?

Theorem

For every Γ , we have $I(\mathcal{A}, \mathcal{B}) \geq 0$, with equality iff \mathcal{A} and \mathcal{B} are statistically independent.

Proof.

Note that $p_i = \sum_j R_{ij}$ and $q_j = \sum_i R_{ij}$. Thus



Lecture 6:
Information
Channel Measures:
 H, I, and C

System Entropies
 Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
Definition
 For BSCs
 Channel Capacity

Epilogue

Theorem

For every Γ , we have $I(\mathcal{A}, \mathcal{B}) \geq 0$, with equality iff \mathcal{A} and \mathcal{B} are statistically independent.

Proof.

Note that $p_i = \sum_j R_{ij}$ and $q_j = \sum_i R_{ij}$. Thus

$$I(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B})$$



Lecture 6:
Information
Channel Measures:
 H, I, and C

System Entropies
 Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
Definition
 For BSCs
 Channel Capacity

Epilogue

Theorem

For every Γ , we have $I(\mathcal{A}, \mathcal{B}) \geq 0$, with equality iff \mathcal{A} and \mathcal{B} are statistically independent.

Proof.

Note that $p_i = \sum_j R_{ij}$ and $q_j = \sum_i R_{ij}$. Thus

$$\begin{aligned} I(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B}) \\ &= \sum_i p_i \log \frac{1}{p_i} + \sum_j q_j \log \frac{1}{q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \end{aligned}$$



Lecture 6:
Information
Channel Measures:
 H, I, and C

System Entropies
 Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
Definition
 For BSCs
 Channel Capacity

Epilogue

Theorem

For every Γ , we have $I(\mathcal{A}, \mathcal{B}) \geq 0$, with equality iff \mathcal{A} and \mathcal{B} are statistically independent.

Proof.

Note that $p_i = \sum_j R_{ij}$ and $q_j = \sum_i R_{ij}$. Thus

$$\begin{aligned} I(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B}) \\ &= \sum_i p_i \log \frac{1}{p_i} + \sum_j q_j \log \frac{1}{q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \\ &= \sum_{i,j} R_{ij} \log \frac{1}{p_i} + \sum_{i,j} R_{ij} \log \frac{1}{q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \end{aligned}$$



Lecture 6:
Information
Channel Measures:
 H, I, and C

System Entropies
 Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
Definition
 For BSCs
 Channel Capacity

Epilogue

Theorem

For every Γ , we have $I(\mathcal{A}, \mathcal{B}) \geq 0$, with equality iff \mathcal{A} and \mathcal{B} are statistically independent.

Proof.

Note that $p_i = \sum_j R_{ij}$ and $q_j = \sum_i R_{ij}$. Thus

$$\begin{aligned}
 I(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B}) \\
 &= \sum_i p_i \log \frac{1}{p_i} + \sum_j q_j \log \frac{1}{q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \\
 &= \sum_{i,j} R_{ij} \log \frac{1}{p_i} + \sum_{i,j} R_{ij} \log \frac{1}{q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \\
 &= \sum_{i,j} R_{ij} \log \frac{1}{p_i q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}}
 \end{aligned}$$



Theorem

For every Γ , we have $I(\mathcal{A}, \mathcal{B}) \geq 0$, with equality iff \mathcal{A} and \mathcal{B} are statistically independent.

Proof.

Note that $p_i = \sum_j R_{ij}$ and $q_j = \sum_i R_{ij}$. Thus

$$\begin{aligned} I(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B}) \\ &= \sum_i p_i \log \frac{1}{p_i} + \sum_j q_j \log \frac{1}{q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \\ &= \sum_{i,j} R_{ij} \log \frac{1}{p_i} + \sum_{i,j} R_{ij} \log \frac{1}{q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \\ &= \sum_{i,j} R_{ij} \log \frac{1}{p_i q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \end{aligned}$$

Note that $\sum_{i,j} R_{ij} = \sum_{i,j} p_i q_j = 1$.



Theorem

For every Γ , we have $I(\mathcal{A}, \mathcal{B}) \geq 0$, with equality iff \mathcal{A} and \mathcal{B} are statistically independent.

Proof.

Note that $p_i = \sum_j R_{ij}$ and $q_j = \sum_i R_{ij}$. Thus

$$\begin{aligned} I(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B}) \\ &= \sum_i p_i \log \frac{1}{p_i} + \sum_j q_j \log \frac{1}{q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \\ &= \sum_{i,j} R_{ij} \log \frac{1}{p_i} + \sum_{i,j} R_{ij} \log \frac{1}{q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \\ &= \sum_{i,j} R_{ij} \log \frac{1}{p_i q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \end{aligned}$$

Note that $\sum_{i,j} R_{ij} = \sum_{i,j} p_i q_j = 1$. Now apply “the corollary” (Gibbs), with distributions (R_{ij}) and $(p_i q_j)$ to get (“=” iff $R_{ij} = p_i q_j$)



Theorem

For every Γ , we have $I(\mathcal{A}, \mathcal{B}) \geq 0$, with equality iff \mathcal{A} and \mathcal{B} are statistically independent.

Proof.

Note that $p_i = \sum_j R_{ij}$ and $q_j = \sum_i R_{ij}$. Thus

$$\begin{aligned} I(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B}) \\ &= \sum_i p_i \log \frac{1}{p_i} + \sum_j q_j \log \frac{1}{q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \\ &= \sum_{i,j} R_{ij} \log \frac{1}{p_i} + \sum_{i,j} R_{ij} \log \frac{1}{q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \\ &= \sum_{i,j} R_{ij} \log \frac{1}{p_i q_j} - \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}} \end{aligned}$$

Note that $\sum_{i,j} R_{ij} = \sum_{i,j} p_i q_j = 1$. Now apply “the corollary” (Gibbs), with distributions (R_{ij}) and $(p_i q_j)$ to get (“=” iff $R_{ij} = p_i q_j$)

$$\sum_{i,j} R_{ij} \log \frac{1}{p_i q_j} \geq \sum_{i,j} R_{ij} \log \frac{1}{R_{ij}}$$



Corollary

For every Γ we have

$$H(\mathcal{A}) \geq H(\mathcal{A} | \mathcal{B}), \quad H(\mathcal{B}) \geq H(\mathcal{B} | \mathcal{A}), \quad H(\mathcal{A}, \mathcal{B}) \leq H(\mathcal{A}) + H(\mathcal{B}).$$

Proof.

We proved earlier that

$$\begin{aligned} I(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) - H(\mathcal{A} | \mathcal{B}) \\ &= H(\mathcal{B}) - H(\mathcal{B} | \mathcal{A}) \\ &= H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B}). \end{aligned}$$

Result follows from the theorem we just proved. □

Lecture 6:
Information
Channel Measures:
 H, I, and C

System Entropies
 Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
Definition
 For BSCs
 Channel Capacity
 Epilogue

Mutual Information in BSCs

We have

$$I(\mathcal{A}, \mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B} \mid \mathcal{A}).$$

Since $H(\mathcal{B}) = H(q)$ and $H(\mathcal{B} \mid \mathcal{A}) = H(P)$, where $q = pP + \bar{p}\bar{P}$,

$$\begin{aligned} I(\mathcal{A}, \mathcal{B}) &= H(q) - H(P) \\ &= H(pP + \bar{p}\bar{P}) - H(P). \end{aligned}$$

Height difference between $H(P)$ and $H(q)$ on graph for H .

Lecture 6:
Information
Channel Measures:
 H, I, and C

System Entropies
 Definitions
 For BSCs

Shannon's First
 Theorem for Γ
 The Idea
 Problems

Mutual Information
 Definition
 For BSCs
 Channel Capacity

Epilogue

Channel Capacity, Definition

Since $I(\mathcal{A}, \mathcal{B})$ represents how much information in \mathcal{A} emerges in \mathcal{B} , when given Γ , we wish to *maximise* $I(\mathcal{A}, \mathcal{B})$ by picking \mathcal{A} suitably (by encoding).

Channel Capacity:

$$C = \max_{\mathcal{A}} I(\mathcal{A}, \mathcal{B})$$

C represents the *maximum amount of information we can send through* Γ . We would *really* like to meet this C in transmission rate.

Lecture 6:
Information
Channel Measures:
H, I, and C

System Entropies
Definitions
For BSCs

Shannon's First
Theorem for Γ
The Idea
Problems

Mutual Information
Definition
For BSCs
Channel Capacity

Epilogue

Channel Capacity, Definition

Since $I(\mathcal{A}, \mathcal{B})$ represents how much information in \mathcal{A} emerges in \mathcal{B} , when given Γ , we wish to *maximise* $I(\mathcal{A}, \mathcal{B})$ by picking \mathcal{A} suitably (by encoding).

Channel Capacity:

$$C = \max_{\mathcal{A}} I(\mathcal{A}, \mathcal{B})$$

C represents the *maximum amount of information we can send through* Γ . We would *really* like to meet this C in transmission rate. **Spoiler**
Shannon's Fundamental Theorem says we can.

Corollary (C is well defined)

$\max_{\mathcal{A}} I(\mathcal{A}, \mathcal{B})$ exists.

Lecture 6:
Information
Channel Measures:
H, I, and C

System Entropies
Definitions
For BSCs

Shannon's First
Theorem for Γ
The Idea
Problems

Mutual Information
Definition
For BSCs
Channel Capacity

Epilogue

Summary

- Introduced to the problem of error-correction. Not a trivial problem.

What we do instead is that we weave this error-correction data into the messages we send. [...] The art is then to weave as little extra information into a message as possible, to achieve as high an error-correction as possible.

- A heap of definitions for Entropy.
- Most importantly, our system entropies lead us to mutual information, which maximum is the channel capacity C ; the best transfer rate we can achieve when sending through Γ .
- We need them in the proof of **Shannon's Fundamental Theorem**, which we will prepare for in the next session.

Lecture 6:
Information
Channel Measures:
H, I, and C

System Entropies
Definitions
For BSCs

Shannon's First
Theorem for Γ
The Idea
Problems

Mutual Information
Definition
For BSCs
Channel Capacity

Epilogue