EE515a – Information Theory II
Winter Quarter 2014

Prof. Jeff Bilmes

University of Washington, Seattle
Department of Electrical Engineering
Winter Quarter, 2014
http://j.ee.washington.edu/~bilmes/classes/ee515a_winter_2014/

Lecture 31 - Feb 24th, 2014
Class Road Map - IT-I

- L19 (1/6): Overview, Communications, Gaussian Channel
- L20 (1/8): Gaussian Channel, band limitation, parallel channels, optimization and duality
- L21 (1/13): parallel channels, colored noise, feedback, matrix inequalities
- L22 (1/15): matrix inequalities, rate distortion.
- – (1/20): Monday holiday
- L23 (1/22): rate distortion for Bernoulli, Gaussian, and Multiple Gaussians with unequal noise
- L24 (1/27): main rate distortion theorem, geometry
- L25 (1/29): computing $R(D)$
- L26 (2/3): computing $R(D)$, alternating minimization
- L27 (2/5): Kolmogorov complexity
- L28 (2/10): algorithmic randomness, universal prob.,
- L29 (2/12): universal compression, LZ compression
- – (2/17): Monday, Holiday
- L30 (2/19): LZ compression, info measures
- L31 (2/24): Info measures
- L32 (2/26): Info measures
- L33 (3/3):
- L34 (3/5):
- L35 (3/10):
- L36 (3/12):


Read Ch. 13 in our book (Cover & Thomas, “Information Theory”).

Read Ch. 14 in our book (Cover & Thomas, “Information Theory”).

Read Ch. 10 in our book (Cover & Thomas, “Information Theory”).

Read Ch. 17 in our book (Cover & Thomas, “Information Theory”) on matrix inequalities.

Read Ch. 9 in our book (Cover & Thomas, “Information Theory”)

Read Ch. 5 in Boyd and Vandenberghe’s Convex Optimization book

Read all readings assigned in EE514a, Fall 2013. (see later lectures on our previous web page (http://j.ee.washington.edu/~bilmes/classes/ee514a_fall_2013/)).
Homework

• No current outstanding HW.
Announcements

- No class Monday, Feb 17th
- Office hours on Mondays, 3:30-4:30.
- As always, email me if you want to skype/google hangout rather than come to office hours, also at different times.
On Final Presentations

- Your task is to give a 10-15 minute presentation that summarizes 2-3 related and significant papers that come from IEEE Transactions on Information Theory (or a very related area).
- The papers must not be ones that we covered in class, although they can be related.
- You need to do the research to find the papers yourself (i.e., that is part of the assignment).
- The majority of the papers must have been published in the last 10 years (so no old or classic papers).
- Your grade will be based on how clear, understandable, and accurate your presentation is (and also milestones).
- This is a real challenge and will require significant work! Many of the papers are complex. To get a good grade, you will need to work very hard to present very complex ideas in an extremely simple yet still precise way.
- Again, don’t expect this to be easy, you might need to try a few topics until you find one that is suitable.
Final Presentation Milestones

All submissions done in PDF file format via our assignment dropbox (https://canvas.uw.edu/courses/880971/assignments)

- **Monday, Feb 24th 11:45pm, tonight:** Updated list of proposed papers decided, based on feedback. Updated writeup with more description. Include PDFs of papers!

- **Monday, March 3rd 11:45pm:** progress report (at most 1 page). Any background papers you needed to read to better understand your core set. Thoughts on coherent and simple unifying presentation.

- **Monday, March 10th, 11:45pm:** updated short (≤ 1 page) writeup on more details of how you will present the ideas in a simple fashion.

- **Final presentations:** Monday, March 17, 2014, 2:30–4:20pm, LOW 102. What to turn in: your slides and a short at most 4 page summary of the papers.
We’ve previously seen that Venn diagrams are a useful way to visualize the relationship between information measures (entropy, etc.).
We’ve previously seen that Venn diagrams are a useful way to visualize the relationship between information measures (entropy, etc.).

But is within these sets? So far, we’ve only said it is “information”
We’ve previously seen that Venn diagrams are a useful way to visualize the relationship between information measures (entropy, etc.).

But is within these sets? So far, we’ve only said it is “information”

We want now to show that set theory and the relation between set theory and information theory can be made more precise in order to:

1. gain intuition
2. help prove theorems
3. lead to new (useful) information theoretic inequalities that are “non-Shannon” (i.e., not previously known).
We’ve previously seen that Venn diagrams are a useful way to visualize the relationship between information measures (entropy, etc.).

But is within these sets? So far, we’ve only said it is “information”

We want now to show that set theory and the relation between set theory and information theory can be made more precise in order to:

1. gain intuition
We’ve previously seen that Venn diagrams are a useful way to visualize the relationship between information measures (entropy, etc.)

But is within these sets? So far, we’ve only said it is “information”

We want now to show that set theory and the relation between set theory and information theory can be made more precise in order to:

1. gain intuition
2. help prove theorems
We’ve previously seen that Venn diagrams are a useful way to visualize the relationship between information measures (entropy, etc.)

But is within these sets? So far, we’ve only said it is “information”

We want now to show that set theory and the relation between set theory and information theory can be made more precise in order to:

1. gain intuition
2. help prove theorems
3. lead to new (useful) information theoretic inequalities that are “non-Shannon” (i.e., not previously known).
We have a set of random variables $X_1, X_2, \ldots, X_n$. 
Definitions: field, atom

- We have a set of random variables $X_1, X_2, \ldots, X_n$.
- For each random variable we associate a set $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$. 
Definitions: field, atom

- We have a set of random variables $X_1, X_2, \ldots, X_n$.
- For each random variable we associate a set $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$.
- A field $\mathcal{F}_n$ can be generated from sets $\tilde{X}_{1:n}$ by taking unions ($\cup$), intersections ($\cap$), complementation ($\tilde{X}^c$), set subtractions/difference ($\setminus$) on combinations of $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$. 

Ex: $n = 2$, 4 such atoms.
Ex: $n = 3$, 8 atoms.
**Definitions: field, atom**

- We have a set of random variables $X_1, X_2, \ldots, X_n$.
- For each random variable we associate a set $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$.
- A field $\mathcal{F}_n$ can be generated from sets $\tilde{X}_1:n$ by taking unions ($\bigcup$), intersections ($\bigcap$), complementation ($\tilde{X}_i^c$), set subtractions/difference ($\setminus$) on combinations of $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$.
- An atom of $\mathcal{F}_n$ are sets of the form

  $$\text{an atom } = \bigcap_{i=1}^{n} Y_i \text{ where } Y_i = \begin{cases} \tilde{X}_i \\ \tilde{X}_i^c \end{cases} \text{ or } \quad (31.1)$$
Definitions: field, atom

- We have a set of random variables $X_1, X_2, \ldots, X_n$.
- For each random variable we associate a set $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$.
- A field $\mathcal{F}_n$ can be generated from sets $\tilde{X}_{1:n}$ by taking unions ($\cup$), intersections ($\cap$), complementation ($\tilde{X}_i^c$), set subtractions/difference ($\setminus$) on combinations of $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$.
- An atom of $\mathcal{F}_n$ are sets of the form

$$\text{an atom} = \bigcap_{i=1}^{n} Y_i \text{ where } Y_i = \begin{cases} \tilde{X}_i \\ \tilde{X}_i^c \end{cases} \text{ or } (31.1)$$

Ex: $n = 2$, 4 such atoms.

Ex: $n = 3$, 8 atoms
In general, there are $|\mathcal{A}| = 2^n$ atoms where $\mathcal{A}$ is the set of atoms.
Info Measures

Definitions: field, atom

- In general, there are $|A| = 2^n$ atoms where $A$ is the set of atoms.
- The atoms are disjoint: Why?
In general, there are $|A| = 2^n$ atoms where $A$ is the set of atoms.
The atoms are disjoint: Why? For any two distinct atoms, there is at least one factor which is a complement.
Definitions: field, atom

- In general, there are $|\mathcal{A}| = 2^n$ atoms where $\mathcal{A}$ is the set of atoms.
- The atoms are disjoint: Why? For any two distinct atoms, there is at least one factor which is a complement.
- There are $2^{2^n}$ elements in the field. Why?
In general, there are $|\mathcal{A}| = 2^n$ atoms where $\mathcal{A}$ is the set of atoms.

The atoms are disjoint: Why? For any two distinct atoms, there is at least one factor which is a complement.

There are $2^{2^n}$ elements in the field. Why? union of disjoint atoms and each atom may be chosen or not chosen.
We will be measuring these sets using a signed measure (meaning it might be positive or negative). In particular, a real-valued function $\mu$ defined on $\mathcal{F}_n$ is called a signed measure if it is set-additive, i.e., for disjoint $A$ and $B$, we have

$$\mu(A \cup B) = \mu(A) + \mu(B)$$ (31.2)
Definitions: signed measure

We will be measuring these sets using a signed measure (meaning it might be positive or negative). In particular, a real-valued function $\mu$ defined on $\mathcal{F}_n$ is called a signed measure if it is set-additive, i.e., for disjoint $A$ and $B$, we have

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

(31.2)

for a signed measure, we must have $\mu(\emptyset) = 0$ since

$$\mu(A) = \mu(A + \emptyset) = \mu(A) + \mu(\emptyset).$$
We will be measuring these sets using a signed measure (meaning it might be positive or negative). In particular, a real-valued function $\mu$ defined on $\mathcal{F}_n$ is called a signed measure if it is set-additive, i.e., for disjoint $A$ and $B$, we have

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

(31.2)

for a signed measure, we must have $\mu(\emptyset) = 0$ since $\mu(A) = \mu(A + \emptyset) = \mu(A) + \mu(\emptyset)$.

Note: For sets $A, B$, set-difference is $A \setminus B \equiv A \cap B^c$. 
We will be measuring these sets using a signed measure (meaning it might be positive or negative). In particular, a real-valued function $\mu$ defined on $\mathcal{F}_n$ is called a signed measure if it is set-additive, i.e., for disjoint $A$ and $B$, we have

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

(31.2)

for a signed measure, we must have $\mu(\emptyset) = 0$ since

$$\mu(A) = \mu(A + \emptyset) = \mu(A) + \mu(\emptyset).$$

Note: For sets $A, B$, set-difference is $A \setminus B \equiv A \cap B^c$.

Thanks to additively, any signed measure on $\mathcal{F}_n$ is defined by its value on the atoms. I.e., any $\tilde{X} \in \mathcal{F}_n$ can be represented as $\tilde{X} = \bigcup_i Y_i$ where $Y_i$ are appropriately chosen atoms.
Definitions: signed measure

- Example: Consider two sets $\tilde{X}_1, \tilde{X}_2$

\[
\mu(\tilde{X}_1) = \mu(\tilde{X}_1 \cap \tilde{X}_c) + \mu(\tilde{X}_1 \cap \tilde{X}_2)
\]
Example: Consider two sets $\tilde{X}_1, \tilde{X}_2$

Signed measure $\mu$ on $\mathcal{F}_n$ determined by four values on four atoms:

$$\mu(\tilde{X}_1 \cap \tilde{X}_2), \mu(\tilde{X}_1^c \cap \tilde{X}_2), \mu(\tilde{X}_1 \cap \tilde{X}_2^c), \mu(\tilde{X}_1^c \cap \tilde{X}_2^c),$$  \hspace{1cm} (31.3)
Definitions: signed measure

Example: Consider two sets $\tilde{X}_1, \tilde{X}_2$

Signed measure $\mu$ on $\mathcal{F}_n$ determined by four values on four atoms:

\[
\mu(\tilde{X}_1 \cap \tilde{X}_2), \mu(\tilde{X}_1^c \cap \tilde{X}_2), \mu(\tilde{X}_1 \cap \tilde{X}_2^c), \mu(\tilde{X}_1^c \cap \tilde{X}_2^c),
\]

For example, we have

\[
\mu(\tilde{X}_1) = \mu\left((\tilde{X}_1 \cap \tilde{X}_2^c) \cup (\tilde{X}_1 \cap \tilde{X}_2)\right) = \mu\left(\tilde{X}_1 \cap \tilde{X}_2^c\right) + \mu\left(\tilde{X}_1 \cap \tilde{X}_2\right)
\]
We said earlier that $\tilde{X}_i$ was associated with a random variable $X_i$. 

Loosely speaking, the set $\tilde{X}_i$ represents the "uncertainty" or "information" contained within $X_i$. 

Ex: Two random variables $X_1, X_2$, define the universal set $\Omega = \tilde{X}_1 \cup \tilde{X}_2$ which is the set of everything (for the current $n=2$) $\Rightarrow X_c \equiv \Omega \setminus X$ for any $X \in F^n$. 

By definition of $\Omega$, one of the atoms is always empty, namely $\tilde{X}_c_1 \cap \tilde{X}_c_2 = (\tilde{X}_1 \cup \tilde{X}_2)_c = \emptyset$, so this atom has no area and is not shown in the Venn diagram previously seen. 

For these two random variables $X_1$ and $X_2$, we already have the Shannon information measures: 

$H(X_1), H(X_2), H(X_1|X_2), H(X_2|X_1), H(X_1,X_2)$.

Lets next associate these with $\mu$ defined on the four atoms.
Random variables

- We said earlier that $\tilde{X}_i$ was associated with a random variable $X_i$.
- Loosely speaking, the set $\tilde{X}_i$ represents the “uncertainty” or “information” contained within $X_i$. 

Ex: Two random variables $X_1, X_2$, define the universal set $\Omega = \tilde{X}_1 \cup \tilde{X}_2$ which is the set of everything (for the current $n = 2$). $X_c \equiv \Omega \setminus X$ for any $X \in F_n$. By definition of $\Omega$, one of the atoms is always empty, namely $\tilde{X}_c_1 \cap \tilde{X}_c_2 = (\tilde{X}_1 \cup \tilde{X}_2)^c = \emptyset$, so this atom has no area and is not shown in the Venn diagram previously seen.

For these two random variables $X_1$ and $X_2$, we already have the Shannon information measures:

- $H(X_1)$,
- $H(X_2)$,
- $H(X_1 | X_2)$,
- $H(X_2 | X_1)$,
- $H(X_1, X_2)$.

Let’s next associate these with $\mu$ defined on the four atoms.
Random variables

- We said earlier that $\tilde{X}_i$ was associated with a random variable $X_i$.
- Loosely speaking, the set $\tilde{X}_i$ represents the “uncertainty” or “information” contained within $X_i$.
- Ex: Two random variables $X_1, X_2$, define the universal set $\Omega = \tilde{X}_1 \cup \tilde{X}_2$ which is the set of everything (for the current $n = 2$) $\Rightarrow X^c \equiv \Omega \setminus X$ for any $X \in \mathcal{F}_n$. 
We said earlier that $\tilde{X}_i$ was associated with a random variable $X_i$.

Loosely speaking, the set $\tilde{X}_i$ represents the “uncertainty” or “information” contained within $X_i$.

**Ex:** Two random variables $X_1, X_2$, define the universal set $\Omega = \tilde{X}_1 \cup \tilde{X}_2$ which is the set of everything (for the current $n = 2$) $\Rightarrow X^c \equiv \Omega \setminus X$ for any $X \in \mathcal{F}_n$.

By definition of $\Omega$, one of the atoms is always empty, namely $\tilde{X}_1^c \cap \tilde{X}_2^c = (\tilde{X}_1 \cup \tilde{X}_2)^c = \emptyset$, so this atom has no area and is not shown in the Venn diagram previously seen.
Random variables

- We said earlier that $\tilde{X}_i$ was associated with a random variable $X_i$.
- Loosely speaking, the set $\tilde{X}_i$ represents the “uncertainty” or “information” contained within $X_i$.
- Ex: Two random variables $X_1, X_2$, define the universal set $\Omega = \tilde{X}_1 \cup \tilde{X}_2$ which is the set of everything (for the current $n = 2$) $\Rightarrow X^c \equiv \Omega \setminus X$ for any $X \in \mathcal{F}_n$.
- By definition of $\Omega$, one of the atoms is always empty, namely $\tilde{X}_1^c \cap \tilde{X}_2^c = (\tilde{X}_1 \cup \tilde{X}_2)^c = \emptyset$, so this atom has no area and is not shown in the Venn diagram previously seen.
- For these two random variables $X_1$ and $X_2$, we already have the Shannon information measures: $H(X_1), H(X_2), H(X_1|X_2), H(X_2|X_1), H(X_1, X_2)$ $I(X_1; X_2)$. 
Random variables

- We said earlier that $\tilde{X}_i$ was associated with a random variable $X_i$.
- Loosely speaking, the set $\tilde{X}_i$ represents the “uncertainty” or “information” contained within $X_i$.
- Ex: Two random variables $X_1, X_2$, define the universal set $\Omega = \tilde{X}_1 \cup \tilde{X}_2$ which is the set of everything (for the current $n = 2$) $\Rightarrow X^c \equiv \Omega \setminus X$ for any $X \in \mathcal{F}_n$.
- By definition of $\Omega$, one of the atoms is always empty, namely $\tilde{X}_1^c \cap \tilde{X}_2^c = (\tilde{X}_1 \cup \tilde{X}_2)^c = \emptyset$, so this atom has no area and is not shown in the Venn diagram previously seen.
- For these two random variables $X_1$ and $X_2$, we already have the Shannon information measures: $H(X_1), H(X_2), H(X_1|X_2), H(X_2|X_1), H(X_1, X_2) I(X_1; X_2)$.
- Let's next associate these with $\mu$ defined on the four atoms.
Signed Measures and Shannon Measures

We can make the following defining associations with signed measure $\mu^*$:

1. $\mu^*(\tilde{X}_1 \cap \tilde{X}_2) = I(X_1; X_2) \quad (31.5)$
2. $\mu^*(\tilde{X}_1 \cap \tilde{X}_2^c) = \mu^*(\tilde{X}_1 \setminus \tilde{X}_2) = H(X_1|X_2) \quad (31.6)$
3. $\mu^*(\tilde{X}_1^c \cap \tilde{X}_2) = \mu^*(\tilde{X}_2 \setminus \tilde{X}_1) = H(X_2|X_1) \quad (31.7)$
4. $\mu^*(\tilde{X}_1^c \cap \tilde{X}_2^c) = \mu^*(\emptyset) = 0 \quad (31.8)$
Signed Measures and Shannon Measures

- We can make the following defining associations with signed measure $\mu^*$:

\[
\begin{align*}
\mu^*(\tilde{X}_1 \cap \tilde{X}_2) &= I(X_1; X_2) \quad (31.5) \\
\mu^*(\tilde{X}_1 \cap \tilde{X}_2^c) &= \mu^*(\tilde{X}_1 \setminus \tilde{X}_2) = H(X_1|X_2) \quad (31.6) \\
\mu^*(\tilde{X}_1^c \cap \tilde{X}_2) &= \mu^*(\tilde{X}_2 \setminus \tilde{X}_1) = H(X_2|X_1) \quad (31.7) \\
\mu^*(\tilde{X}_1^c \cap \tilde{X}_2^c) &= \mu^*(\emptyset) = 0 \quad (31.8)
\end{align*}
\]

- We have instantiated the measures of the four atoms with values (could be arbitrary values, but we chose to use entropic quantities).
Signed Measures and Shannon Measures

- We can make the following defining associations with signed measure $\mu^*$:

  \[
  \mu^*(\tilde{X}_1 \cap \tilde{X}_2) = I(X_1; X_2) \quad (31.5)
  \]

  \[
  \mu^*(\tilde{X}_1 \cap \tilde{X}_2^c) = \mu^*(\tilde{X}_1 \setminus \tilde{X}_2) = H(X_1|X_2) \quad (31.6)
  \]

  \[
  \mu^*(\tilde{X}_1^c \cap \tilde{X}_2) = \mu^*(\tilde{X}_2 \setminus \tilde{X}_1) = H(X_2|X_1) \quad (31.7)
  \]

  \[
  \mu^*(\tilde{X}_1^c \cap \tilde{X}_2^c) = \mu^*(\emptyset) = 0 \quad (31.8)
  \]

- We have instantiated the measures of the four atoms with values (could be arbitrary values, but we chose to use entropic quantities).

- Given these definitions, what would $\mu^*(\tilde{X}_1)$, $\mu^*(\tilde{X}_2)$, and $\mu^*(\tilde{X}_1 \cup \tilde{X}_2)$ be?
Signed Measures and Shannon Measures

- We can make the following defining associations with signed measure $\mu^*$:

$$
\begin{align*}
\mu^*(\tilde{X}_1 \cap \tilde{X}_2) &= I(X_1; X_2) \quad (31.5) \\
\mu^*(\tilde{X}_1 \cap \tilde{X}_2^c) &= \mu^*(\tilde{X}_1 \setminus \tilde{X}_2) = H(X_1|X_2) \quad (31.6) \\
\mu^*(\tilde{X}_1^c \cap \tilde{X}_2) &= \mu^*(\tilde{X}_2 \setminus \tilde{X}_1) = H(X_2|X_1) \quad (31.7) \\
\mu^*(\tilde{X}_1^c \cap \tilde{X}_2^c) &= \mu^*(\emptyset) = 0 \quad (31.8)
\end{align*}
$$

- We have instantiated the measures of the four atoms with values (could be arbitrary values, but we chose to use entropic quantities).

- Given these definitions, what would $\mu^*(\tilde{X}_1), \mu^*(\tilde{X}_2)$, and $\mu^*(\tilde{X}_1 \cup \tilde{X}_2)$ be?

$$
\mu^*(\tilde{X}_1)
$$
Signed Measures and Shannon Measures

- We can make the following defining associations with signed measure $\mu^*$:

$$
\mu^*(\tilde{X}_1 \cap \tilde{X}_2) = I(X_1; X_2) \quad (31.5)
$$

$$
\mu^*(\tilde{X}_1 \cap \tilde{X}_2^c) = \mu^*(\tilde{X}_1 \setminus \tilde{X}_2) = H(X_1|X_2) \quad (31.6)
$$

$$
\mu^*(\tilde{X}_1^c \cap \tilde{X}_2) = \mu^*(\tilde{X}_2 \setminus \tilde{X}_1) = H(X_2|X_1) \quad (31.7)
$$

$$
\mu^*(\tilde{X}_1^c \cap \tilde{X}_2^c) = \mu^*(\emptyset) = 0 \quad (31.8)
$$

- We have instantiated the measures of the four atoms with values (could be arbitrary values, but we chose to use entropic quantities).

- Given these definitions, what would $\mu^*(\tilde{X}_1)$, $\mu^*(\tilde{X}_2)$, and $\mu^*(\tilde{X}_1 \cup \tilde{X}_2)$ be?

$$
\mu^*(\tilde{X}_1) = \mu^*((\tilde{X}_1 \cap \tilde{X}_2) \cup (\tilde{X}_1 \cap \tilde{X}_2^c)) \quad (31.9)
$$
Signed Measures and Shannon Measures

- We can make the following defining associations with signed measure $\mu^*$:

$$\mu^*(\tilde{X}_1 \cap \tilde{X}_2) = I(X_1; X_2) \quad (31.5)$$

$$\mu^*(\tilde{X}_1 \cap \tilde{X}_2^c) = \mu^*(\tilde{X}_1 \setminus \tilde{X}_2) = H(X_1|X_2) \quad (31.6)$$

$$\mu^*(\tilde{X}_1^c \cap \tilde{X}_2) = \mu^*(\tilde{X}_2 \setminus \tilde{X}_1) = H(X_2|X_1) \quad (31.7)$$

$$\mu^*(\tilde{X}_1^c \cap \tilde{X}_2^c) = \mu^*(\emptyset) = 0 \quad (31.8)$$

- We have instantiated the measures of the four atoms with values (could be arbitrary values, but we chose to use entropic quantities).

- Given these definitions, what would $\mu^*(\tilde{X}_1)$, $\mu^*(\tilde{X}_2)$, and $\mu^*(\tilde{X}_1 \cup \tilde{X}_2)$ be?

$$\mu^*(\tilde{X}_1) = \mu^*((\tilde{X}_1 \cap \tilde{X}_2) \cup (\tilde{X}_1 \cap \tilde{X}_2^c)) \quad (31.9)$$

$$= I(X_1; X_2) + H(X_1|X_2) = H(X_1) \quad (31.10)$$
What would $\mu^*(\tilde{X}_2)$ be?

\[ \mu^*(\tilde{X}_2) = H(X_2) \] (31.11)

What about $\mu^*(\tilde{X}_1 \cup \tilde{X}_2)$?

\[ \mu^*(\tilde{X}_1 \cup \tilde{X}_2) = \mu(\bigcup A) \] (31.12)

\[ = I(X_1; X_2) + H(X_1 | X_2) + H(X_2 | X_1) + 0 \] (31.13)

\[ = H(X_1, X_2) \] (31.14)

So, we have defined $\mu^*$ only on the atoms, and from this we have, using the signed measure property and set theory, fully recovered all the rest of the Shannon information values.
What would $\mu^*(\tilde{X}_2)$ be?

$$\mu^*(\tilde{X}_2) = H(X_2) \tag{31.11}$$
Signed Measures and Shannon Measures

- What would $\mu^*(\tilde{X}_2)$ be?

$$
\mu^*(\tilde{X}_2) = H(X_2)
$$

(31.11)

- What about $\mu^*(\tilde{X}_1 \cup \tilde{X}_2)$?

$$
(31.14)$$
Signed Measures and Shannon Measures

What would $\mu^*(\tilde{X}_2)$ be?

$$\mu^*(\tilde{X}_2) = H(X_2) \quad (31.11)$$

What about $\mu^*(\tilde{X}_1 \cup \tilde{X}_2)$?

$$\mu^*(\tilde{X}_1 \cup \tilde{X}_2) \quad (31.14)$$
Signed Measures and Shannon Measures

What would $\mu^*(\tilde{X}_2)$ be?

$$\mu^*(\tilde{X}_2) = H(X_2)$$  \hspace{1cm} (31.11)

What about $\mu^*(\tilde{X}_1 \cup \tilde{X}_2)$?

$$\mu^*(\tilde{X}_1 \cup \tilde{X}_2) = \mu \left( \bigcup \text{ all atoms } A \right)$$ \hspace{1cm} (31.12)

So, we have defined $\mu^*$ only on the atoms, and from this we have, using the signed measure property and set theory, fully recovered all the rest of the Shannon information values.

$$= H(X_1, X_2)$$ \hspace{1cm} (31.14)
Signed Measures and Shannon Measures

- What would $\mu^*(\tilde{X}_2)$ be?
  \[
  \mu^*(\tilde{X}_2) = H(X_2)
  \]  
  (31.11)

- What about $\mu^*(\tilde{X}_1 \cup \tilde{X}_2)$?
  \[
  \mu^*(\tilde{X}_1 \cup \tilde{X}_2) = \mu \left( \bigcup_{\text{all atoms } A} A \right)
  \]
  \[
  = I(X_1; X_2) + H(X_1|X_2) + H(X_2|X_1) + 0
  \]  
  (31.13)
What would $\mu^*(\tilde{X}_2)$ be?

$$\mu^*(\tilde{X}_2) = H(X_2) \quad (31.11)$$

What about $\mu^*(\tilde{X}_1 \cup \tilde{X}_2)$?

$$\mu^*(\tilde{X}_1 \cup \tilde{X}_2) = \mu \left( \bigcup_{\text{all atoms } A} A \right) \quad (31.12)$$

$$= I(X_1; X_2) + H(X_1|X_2) + H(X_2|X_1) + 0 \quad (31.13)$$

$$= H(X_1, X_2) \quad (31.14)$$
Signed Measures and Shannon Measures

- What would $\mu^* (\tilde{X}_2)$ be?

$$
\mu^* (\tilde{X}_2) = H(X_2) \quad (31.11)
$$

- What about $\mu^* (\tilde{X}_1 \cup \tilde{X}_2)$?

$$
\mu^* (\tilde{X}_1 \cup \tilde{X}_2) = \mu \left( \bigcup_{\text{all atoms } A} A \right) \quad (31.12)
$$

$$
= I(X_1; X_2) + H(X_1|X_2) + H(X_2|X_1) + 0 \quad (31.13)
$$

$$
= H(X_1, X_2) \quad (31.14)
$$

- So, we have defined $\mu^*$ only on the atoms, and from this we have, using the signed measure property and set theory, fully recovered all the rest of the Shannon information values.
What if we define $\mu^*$ only on the unions of sets. I.e., we make the following defining associations:

$$\mu^*(\emptyset) = 0$$  \hspace{1cm} (31.15)

$$\mu^*(\tilde{X}_1) = H(X_1)$$  \hspace{1cm} (31.16)

$$\mu^*(\tilde{X}_2) = H(X_2)$$  \hspace{1cm} (31.17)

$$\mu^*(\tilde{X}_1 \cup \tilde{X}_2) = H(X_1, X_2)$$  \hspace{1cm} (31.18)
Unions of sets

- What if we define $\mu^*$ only on the unions of sets. I.e., we make the following defining associations:

  \[
  \begin{align*}
  \mu^*(\emptyset) &= 0 \\
  \mu^*(\tilde{X}_1) &= H(X_1) \\
  \mu^*(\tilde{X}_2) &= H(X_2) \\
  \mu^*(\tilde{X}_1 \cup \tilde{X}_2) &= H(X_1, X_2)
  \end{align*}
  \]

- Then from this, we can (using set theory) get the rest of the values, $I(X_1; X_2)$, $H(X_1 | X_2)$, $H(X_2 | X_1)$. 

E.g., we get:

\[
\begin{align*}
\mu^*(\tilde{X}_1 \cap \tilde{X}_2) &= \mu^*(\tilde{X}_1) + \mu^*(\tilde{X}_2) - \mu^*(\tilde{X}_1 \cup \tilde{X}_2) \\
&= H(X_1) + H(X_2) - H(X_1, X_2) = I(X_1; X_2)
\end{align*}
\]
Unions of sets

- What if we define $\mu^*$ only on the unions of sets. I.e., we make the following defining associations:

$$
\mu^*(\emptyset) = 0 \quad (31.15)
$$

$$
\mu^*(\tilde{X}_1) = H(X_1) \quad (31.16)
$$

$$
\mu^*(\tilde{X}_2) = H(X_2) \quad (31.17)
$$

$$
\mu^*(\tilde{X}_1 \cup \tilde{X}_2) = H(X_1, X_2) \quad (31.18)
$$

- Then from this, we can (using set theory) get the rest of the values, $I(X_1; X_2)$, $H(X_1|X_2)$, $H(X_2|X_1)$.

- E.g., we get:

$$
\mu(\tilde{X}_1 \cap \tilde{X}_2) = \mu(\tilde{X}_1) + \mu(\tilde{X}_2) - \mu(\tilde{X}_1 \cup \tilde{X}_2) \quad (31.19)
$$

$$
= H(X_1) + H(X_2) - H(X_1, X_2) = I(X_1; X_2) \quad (31.20)
$$
So we have recovered Shannon’s information measures with the following correspondence:

\[ H/I \leftrightarrow \mu^* \]  \hspace{1cm} (31.21)

\[ , \leftrightarrow \cup \]  \hspace{1cm} (31.22)

\[ ; \leftrightarrow \cap \]  \hspace{1cm} (31.23)

\[ | \leftrightarrow \setminus \quad //\text{set minus} \]  \hspace{1cm} (31.24)
So we have recovered Shannon’s information measures with the following correspondence:

\[ H/I \leftrightarrow \mu^* \quad (31.21) \]
\[ , \leftrightarrow \bigcup \quad (31.22) \]
\[ ; \leftrightarrow \bigcap \quad (31.23) \]
\[ | \leftrightarrow \setminus \quad //\text{set minus} \quad (31.24) \]

Note: with measure notation, no distinction between \( H \) and \( I \), we could identify \( H(X; Y) = I(X; Y) \), where \( H(X; Y) \) vs. \( H(X, Y) \) would be distinguished only by a semicolon rather than a comma.
So we have recovered Shannon’s information measures with the following correspondence:

\[ H/I \leftrightarrow \mu^* \quad (31.21) \]
\[ , \leftrightarrow \cup \quad (31.22) \]
\[ ; \leftrightarrow \cap \quad (31.23) \]
\[ | \leftrightarrow \setminus \quad \text{/set minus} \quad (31.24) \]

Note: with measure notation, no distinction between \( H \) and \( I \), we could identify \( H(X;Y) = I(X;Y) \), where \( H(X;Y) \) vs. \( H(X,Y) \) would be distinguished only by a semicolon rather than a comma. but we don’t do that.
Recovering Shannon

- So we have recovered Shannon’s information measures with the following correspondence:

\[ H/I \leftrightarrow \mu^* \quad (31.21) \]
\[ , \leftrightarrow \cup \quad (31.22) \]
\[ ; \leftrightarrow \cap \quad (31.23) \]
\[ | \leftrightarrow \setminus \quad \text{//set minus} \quad (31.24) \]

- Note: with measure notation, no distinction between \( H \) and \( I \), we could identify \( H(X;Y) = I(X;Y) \), where \( H(X;Y) \) vs. \( H(X,Y) \) would be distinguished only by a semicolon rather than a comma. but we don’t do that.

- Does this generalize? The hope is that if there is an information theoretic identity, it would occur iff there is a set theory identity.
For example, a simple example of a well-known property in set theory: The inclusion-exclusion formula for measures, is as follows:

\[ \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \]  
(31.25)

which follows since

\[ \mu(A \cup B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) + \mu(A \cap B) - \mu(A \cap B) \]  
(31.28)
For example, a simple example of a well-known property in set theory: The inclusion-exclusion formula for measures, is as follows:

\[
\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)
\]  

which follows since

\[
\mu(A \cup B)
\]  

(31.25)
For example, a simple example of a well-known property in set theory: The inclusion-exclusion formula for measures, is as follows:

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \quad (31.25)$$

which follows since

$$\mu(A \cup B) = \mu(A \cup (B \setminus A)) \quad (31.28)$$
For example, a simple example of a well-known property in set theory: The inclusion-exclusion formula for measures, is as follows:

\[ \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \]  \hspace{1cm} (31.25)

which follows since

\[ \mu(A \cup B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) \]  \hspace{1cm} (31.26)

\[ \mu(A \cup B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) \]  \hspace{1cm} (31.27)

Equating our measures, we see that the entropy/mutual information formula is just inclusion-exclusion:

\[ H(X_1, X_2) = H(X_1) + H(X_2) - I(X_1; X_2) \]  \hspace{1cm} (31.29)
Recovering Shannon

For example, a simple example of a well-known property in set theory: The inclusion-exclusion formula for measures, is as follows:

\[ \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \quad (31.25) \]

which follows since

\[ \mu(A \cup B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) \quad (31.26) \]
\[ = \mu(A) + \mu(B \setminus A) + \mu(A \cap B) - \mu(A \cap B) \quad (31.27) \]
\[ = \mu(A) + \mu(B \setminus A) + \mu(A \cap B) - \mu(A \cap B) \quad (31.28) \]
For example, a simple example of a well-known property in set theory: The inclusion-exclusion formula for measures, is as follows:

\[ \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \]  

(31.25)

which follows since

\[ \mu(A \cup B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) \]  

(31.26)

\[ = \mu(A) + \mu(B \setminus A) + \mu(A \cap B) - \mu(A \cap B) \]  

(31.27)

\[ = \mu(A) + \mu(B) - \mu(A \cap B) \]  

(31.28)
For example, a simple example of a well-known property in set theory: The inclusion-exclusion formula for measures, is as follows:

\[ \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \quad (31.25) \]

which follows since

\[
\begin{align*}
\mu(A \cup B) &= \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) \\
&= \mu(A) + \mu(B \setminus A) + \mu(A \cap B) - \mu(A \cap B) \\
&= \mu(A) + \mu(B) - \mu(A \cap B)
\end{align*}
\quad (31.26) \]

Equating our measures, we see that the entropy/mutual information formula is just inclusion-exclusion:

\[ H(X_1, X_2) = H(X_1) + H(X_2) - I(X_1; X_2) \quad (31.29) \]
General case, \( n \geq 2 \)

- \( n \) random variables \( X_1, \ldots, X_n \) corresponding to sets \( \tilde{X}_i \), and with \( [n] = \{1, 2, \ldots, n\} \) the index set.
General case, $n \geq 2$

- $n$ random variables $X_1, \ldots, X_n$ corresponding to sets $\tilde{X}_i$, and with $[n] = \{1, 2, \ldots, n\}$ the index set.
- $\Omega = \bigcup_i \tilde{X}_i$ is the universe, and empty atom again is:

$$A_0 = \bigcap_{i \in [n]} \tilde{X}_i^c = \left(\bigcup_i \tilde{X}_i\right)^c = \emptyset$$  \hspace{1cm} (31.30)
General case, $n \geq 2$

- $n$ random variables $X_1, \ldots, X_n$ corresponding to sets $\tilde{X}_i$, and with $[n] = \{1, 2, \ldots, n\}$ the index set.
- $\Omega = \bigcup_i \tilde{X}_i$ is the universe, and empty atom again is:
  $$A_0 = \bigcap_{i \in [n]} \tilde{X}_i^c = \left(\bigcup_i \tilde{X}_i\right)^c = \emptyset$$ (31.30)

- Non-empty atoms are $\mathcal{A} \triangleq \{\text{all atoms}\} \setminus \{A_0\}$ which are those that are not assuredly empty, so $|\mathcal{A}| = 2^n - 1$. 

Notation: for $G \subseteq [n]$, $X_G = (X_i, i \in G)$ for index set $G$.

Notation: for $G \subseteq [n]$, $\tilde{X}_G = \bigcup_{i \in G} \tilde{X}_i$. 

Definition, non-empty unions (note strictness on left side):
$$B \triangleq \{\tilde{X}_G : \emptyset \subset G \subseteq [n]\}$$ (31.31)
General case, $n \geq 2$

- $n$ random variables $X_1, \ldots, X_n$ corresponding to sets $\tilde{X}_i$, and with $[n] = \{1, 2, \ldots, n\}$ the index set.
- $\Omega = \bigcup_i \tilde{X}_i$ is the universe, and empty atom again is:
  
  $$A_0 = \bigcap_{i \in [n]} \tilde{X}_i^c = \left(\bigcup_i \tilde{X}_i\right)^c = \emptyset$$  

  \[ (31.30) \]

- Non-empty atoms are $\mathcal{A} \triangleq \{\text{all atoms}\} \setminus \{A_0\}$ which are those that are not assuredly empty, so $|\mathcal{A}| = 2^n - 1$.
- When values of $\mu(\cdot)$ is given for all $\mathcal{A}$, then this defines $\mu(\cdot)$ on all $\mathcal{F}_n$. Why?
General case, $n \geq 2$

- $n$ random variables $X_1, \ldots, X_n$ corresponding to sets $\tilde{X}_i$, and with $[n] = \{1, 2, \ldots, n\}$ the index set.
- $\Omega = \bigcup_i \tilde{X}_i$ is the universe, and empty atom again is:

$$A_0 = \bigcap_{i \in [n]} \tilde{X}_i^c = \left(\bigcup_i \tilde{X}_i\right)^c = \emptyset$$  \hspace{1cm} (31.30)

- Non-empty atoms are $\mathcal{A} \triangleq \{\text{all atoms}\} \setminus \{A_0\}$ which are those that are not assuredly empty, so $|\mathcal{A}| = 2^n - 1$.
- When values of $\mu(\cdot)$ is given for all $\mathcal{A}$, then this defines $\mu(\cdot)$ on all $\mathcal{F}_n$. Why? Since the atoms $\mathcal{A}$ partition and hence fully cover $\Omega$. 

Note: For $G \subseteq [n]$, $X_G = (X_i, i \in G)$ for index set $G$. Notation: for $G \subseteq [n]$, $\tilde{X}_G = \bigcup_{i \in G} \tilde{X}_i$. Definition, non-empty unions (note strictness on left side): $B \triangleq \{\tilde{X}_G: \emptyset \subset G \subseteq [n]\}$ (31.31)
General case, \( n \geq 2 \)

- \( n \) random variables \( X_1, \ldots, X_n \) corresponding to sets \( \tilde{X}_i \), and with \([n] = \{1, 2, \ldots, n\}\) the index set.
- \( \Omega = \bigcup_i \tilde{X}_i \) is the universe, and empty atom again is:

\[
A_0 = \bigcap_{i \in [n]} \tilde{X}_i^c = \left( \bigcup_i \tilde{X}_i \right)^c = \emptyset \tag{31.30}
\]

Non-empty atoms are \( \mathcal{A} \triangleq \{\text{all atoms}\} \setminus \{A_0\} \), which are those that are not assuredly empty, so \(|\mathcal{A}| = 2^n - 1\).

- When values of \( \mu(\cdot) \) is given for all \( \mathcal{A} \), then this defines \( \mu(\cdot) \) on all \( \mathcal{F}_n \). Why? Since the atoms \( \mathcal{A} \) partition and hence fully cover \( \Omega \).
- Notation: for \( G \subseteq [n] \), \( X_G = (X_i, i \in G) \) for index set \( G \).
General case, \( n \geq 2 \)

- \( n \) random variables \( X_1, \ldots, X_n \) corresponding to sets \( \tilde{X}_i \), and with 
  \([n] = \{1, 2, \ldots, n\}\) the index set.
- \( \Omega = \bigcup_i \tilde{X}_i \) is the universe, and empty atom again is:
  \[
  A_0 = \bigcap_{i \in [n]} \tilde{X}_i^c = \left( \bigcup_i \tilde{X}_i \right)^c = \emptyset \quad (31.30)
  \]
- Non-empty atoms are \( A \triangleq \{\text{all atoms}\} \setminus \{A_0\} \) which are those that are not assuredly empty, so \(|A| = 2^n - 1\).
- When values of \( \mu(\cdot) \) is given for all \( A \), then this defines \( \mu(\cdot) \) on all \( \mathcal{F}_n \). Why? Since the atoms \( A \) partition and hence fully cover \( \Omega \).
- Notation: for \( G \subseteq [n], X_G = (X_i, i \in G) \) for index set \( G \).
- Notation: for \( G \subseteq [n], \tilde{X}_G = \bigcup_{i \in G} \tilde{X}_i \).
General case, \( n \geq 2 \)

- \( n \) random variables \( X_1, \ldots, X_n \) corresponding to sets \( \tilde{X}_i \), and with
  \( [n] = \{1, 2, \ldots, n\} \) the index set.
- \( \Omega = \bigcup_i \tilde{X}_i \) is the universe, and empty atom again is:

\[
A_0 = \bigcap_{i \in [n]} \tilde{X}_i^c = \left( \bigcup_i \tilde{X}_i \right)^c = \emptyset \quad (31.30)
\]

- Non-empty atoms are \( \mathcal{A} \triangleq \{ \text{all atoms} \} \setminus \{ A_0 \} \) which are those that are not assuredly empty, so \( |\mathcal{A}| = 2^n - 1 \).
- When values of \( \mu(\cdot) \) is given for all \( \mathcal{A} \), then this defines \( \mu(\cdot) \) on all \( \mathcal{F}_n \). Why? Since the atoms \( \mathcal{A} \) partition and hence fully cover \( \Omega \).
- Notation: for \( G \subseteq [n] \), \( X_G = (X_i, i \in G) \) for index set \( G \).
- Notation: for \( G \subseteq [n] \), \( \tilde{X}_G = \bigcup_{i \in G} \tilde{X}_i \).
- Definition, non-empty unions (note strictness on left side):

\[
\mathcal{B} \triangleq \left\{ \tilde{X}_G : \emptyset \subset G \subseteq [n] \right\} \quad (31.31)
\]
What needs to be specified: getting atoms from sets

**Theorem 31.3.1**

Signed measure \( \mu \) on \( \mathcal{F}_n \) is fully specified by \( \{ \mu(B) : B \in \mathcal{B} \} \) which can be any set of real numbers.

- So before, we defined \( \mu \) on all atoms and noted that this allowed us to compute \( \mu \) everywhere else.
What needs to be specified: getting atoms from sets

Theorem 31.3.1

Signed measure $\mu$ on $\mathcal{F}_n$ is fully specified by $\{\mu(B) : B \in \mathcal{B}\}$ which can be any set of real numbers.

- So before, we defined $\mu$ on all atoms and noted that this allowed us to compute $\mu$ everywhere else.
- Here, in the above theorem, we are defining $\mu$ only on elements of $\mathcal{B}$ and are again saying that this allows us to compute the values of $\mu$ everywhere else.
What needs to be specified: getting atoms from sets

**Theorem 31.3.1**

Signed measure $\mu$ on $\mathcal{F}_n$ is fully specified by $\{\mu(B) : B \in \mathcal{B}\}$ which can be any set of real numbers.

- So before, we defined $\mu$ on all atoms and noted that this allowed us to compute $\mu$ everywhere else.
- Here, in the above theorem, we are defining $\mu$ only on elements of $\mathcal{B}$ and are again saying that this allows us to compute the values of $\mu$ everywhere else.
- We will see that this allows us to generate all standard mutual-information quantities, and some other less standard ones, using just entropy to fill out $\mu$ on $\mathcal{B}$. 
Consider a simple signed measure $\mu(A) = |A|$, the cardinality (or size or counting) measure (but the same idea works for any signed measure)
Consider a simple signed measure $\mu(A) = |A|$, the cardinality (or size or counting) measure (but the same idea works for any signed measure)

First note, from the binomial expansion:

$$0$$
Consider a simple signed measure $\mu(A) = |A|$, the cardinality (or size or counting) measure (but the same idea works for any signed measure).

First note, from the binomial expansion:

$$0 = 1 - 1$$
Consider a simple signed measure $\mu(A) = |A|$, the cardinality (or size or counting) measure (but the same idea works for any signed measure).

First note, from the binomial expansion:

$$0 = 1 - 1 = (1 - 1)^n$$

(31.32)
Consider a simple signed measure \( \mu(A) = |A| \), the cardinality (or size or counting) measure (but the same idea works for any signed measure)

First note, from the binomial expansion:

\[
0 = 1 - 1 = (1 - 1)^n = \sum_{\ell=0}^{n} \binom{n}{\ell} (-1)^\ell (1)^{n-\ell}
\]
Consider a simple signed measure $\mu(A) = |A|$, the cardinality (or size or counting) measure (but the same idea works for any signed measure).

First note, from the binomial expansion:

\[
0 = 1 - 1 = (1 - 1)^n = \sum_{\ell=0}^{n} \binom{n}{\ell} (-1)^\ell (1)^{n-\ell} = 1 - \binom{n}{1} + \binom{n}{2} - \ldots + (-1)^n \binom{n}{n}
\]
Consider a simple signed measure $\mu(A) = |A|$, the cardinality (or size or counting) measure (but the same idea works for any signed measure).

First note, from the binomial expansion:

$$0 = 1 - 1 = (1 - 1)^n$$

$$= \sum_{\ell=0}^{n} \binom{n}{\ell} (-1)^\ell (1)^{n-\ell}$$

$$= 1 - \binom{n}{1} + \binom{n}{2} - \ldots + (-1)^n \binom{n}{n}$$

We have $n$ sets $A_i$, for $i = 1 \ldots n$ such that $A_i \subseteq \Omega$. 

We have $n$ sets $A_i$, for $i = 1 \ldots n$ such that $A_i \subseteq \Omega$. What we wish to prove is the form of the inclusion/exclusion formula.

\[
|\bigcap_{i=1}^{n} A_i| = \sum_{i=1}^{n} |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cup A_j| + \sum_{1 \leq i < j < k \leq n} |A_i \cup A_j \cup A_k| + \ldots + (-1)^{n-1} |A_1 \cup A_2 \cup \ldots \cup A_n|
\] (31.35)

(31.36)

(31.37)
We have \( n \) sets \( A_i \), for \( i = 1 \ldots n \) such that \( A_i \subseteq \Omega \). What we wish to prove is the form of the inclusion/exclusion formula.

\[
| \cap_{i=1}^{n} A_i | = \sum_{i=1}^{n} |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cup A_j| + \sum_{1 \leq i < j < k \leq n} |A_i \cup A_j \cup A_k| + \ldots + (-1)^{n-1}|A_1 \cup A_2 \cup \ldots \cup A_n| \tag{31.35}
\]

Note the pattern: first we over count, then we undercount, and then overcount, etc. . . until the last term finally fixes things.
We have \( n \) sets \( A_i \), for \( i = 1 \ldots n \) such that \( A_i \subseteq \Omega \). What we wish to prove is the form of the inclusion/exclusion formula.

\[
| \cap_{i=1}^n A_i | = \sum_{i=1}^n |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cup A_j| \\
+ \sum_{1 \leq i < j < k \leq n} |A_i \cup A_j \cup A_k| + \ldots \\
+ (-1)^{n-1}|A_1 \cup A_2 \cup \ldots \cup A_n|
\]  

(31.35) (31.36) (31.37)

Note the pattern: first we over count, then we undercount, and then overcount, etc. \ldots until the last term finally fixes things.

Special case of sieve methods: general mathematical methods to count sizes of sets of integers (e.g., sieve of Eratosthenes).
Consider an \( x \in \Omega \) where \( x \in A_i \) for all \( i = 1 \ldots n \).
Consider an $x \in \Omega$ where $x \in A_i$ for all $i = 1 \ldots n$.

Then the l.h.s. of Equation 31.37 contributes only 1 (unity) for this $x$. 
Consider an $x \in \Omega$ where $x \in A_i$ for all $i = 1 \ldots n$.

Then the l.h.s. of Equation 31.37 contributes only 1 (unity) for this $x$.

For the r.h.s. of Equation 31.37, we want to look what the contribution for this particular $x$ will be, and we do this on the next slide.
For the r.h.s. of Equation 31.37, we want to look what the contribution for this particular $x$ will be, and it is:

$$n - \binom{n}{2} + \binom{n}{3} - \binom{n}{4} + \ldots + (-1)^{n-1} \binom{n}{n}$$

(31.38)

Thus, the contribution for any particular $x$ is one both on the l.h.s. and on the r.h.s. of the equation.
For the r.h.s. of Equation 31.37, we want to look what the contribution for this particular \( x \) will be, and it is:

\[
\begin{align*}
\binom{n}{2} - \binom{n}{3} + \binom{n}{4} - \ldots - (-1)^{n-1} \binom{n}{n} \\
&= (-1)^0 \binom{n}{1} + (-1)^1 \binom{n}{2} + (-1)^2 \binom{n}{3} + (-1)^3 \binom{n}{4} + \ldots + (-1)^{n-1} \binom{n}{n}
\end{align*}
\]  

(31.38)

(31.39)

Thus, the contribution for any particular \( x \) is one both on the l.h.s. and on the r.h.s. of the equation.
For the r.h.s. of Equation 31.37, we want to look what the contribution for this particular $x$ will be, and it is:

$$n - \binom{n}{2} + \binom{n}{3} - \binom{n}{4} + \ldots + (-1)^{n-1} \binom{n}{n}$$  \hspace{1cm} (31.38)

$$= (-1)^0 \binom{n}{1} + (-1)^1 \binom{n}{2} + (-1)^2 \binom{n}{3} + (-1)^3 \binom{n}{4} + \ldots + (-1)^{n-1} \binom{n}{n}$$  \hspace{1cm} (31.39)

$$= (-1) \left[ (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + (-1)^3 \binom{n}{3} + (-1)^4 \binom{n}{4} + \ldots + (-1)^n \binom{n}{n} \right]$$  \hspace{1cm} (31.40)

Thus, the contribution for any particular $x$ is one both on the l.h.s. and on the r.h.s. of the equation.
For the r.h.s. of Equation 31.37, we want to look what the contribution for this particular \( x \) will be, and it is:

\[
\begin{align*}
&= n \left( \begin{array}{c} n \\ 2 \end{array} \right) + \left( \begin{array}{c} n \\ 3 \end{array} \right) - \left( \begin{array}{c} n \\ 4 \end{array} \right) + \ldots + (-1)^{n-1} \left( \begin{array}{c} n \\ n \end{array} \right) \\
&= (-1)^0 \left( \begin{array}{c} n \\ 1 \end{array} \right) + (-1)^1 \left( \begin{array}{c} n \\ 2 \end{array} \right) + (-1)^2 \left( \begin{array}{c} n \\ 3 \end{array} \right) + (-1)^3 \left( \begin{array}{c} n \\ 4 \end{array} \right) + \ldots + (-1)^{n-1} \left( \begin{array}{c} n \\ n \end{array} \right) \\
&= (-1)^0 \left( \begin{array}{c} n \\ 1 \end{array} \right) + (-1)^1 \left( \begin{array}{c} n \\ 2 \end{array} \right) + (-1)^2 \left( \begin{array}{c} n \\ 3 \end{array} \right) + (-1)^3 \left( \begin{array}{c} n \\ 4 \end{array} \right) + \ldots + (-1)^{n} \left( \begin{array}{c} n \\ n \end{array} \right)
\end{align*}
\]

Thus, the contribution for any particular \( x \) is one both on the l.h.s. and on the r.h.s. of the equation.
For the r.h.s. of Equation 31.37, we want to look what the contribution for this particular $x$ will be, and it is:

$$n - \binom{n}{2} + \binom{n}{3} - \binom{n}{4} + \ldots + (-1)^{n-1} \binom{n}{n}$$  \hspace{1cm} (31.38)$$

$$= (-1)^0 \binom{n}{1} + (-1)^1 \binom{n}{2} + (-1)^2 \binom{n}{3} + (-1)^3 \binom{n}{4} + \ldots + (-1)^{n-1} \binom{n}{n}$$  \hspace{1cm} (31.39)$$

$$= (-1) \left[ (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + (-1)^3 \binom{n}{3} + (-1)^4 \binom{n}{4} + \ldots + (-1)^n \binom{n}{n} \right]$$  \hspace{1cm} (31.40)$$

$$= (-1) \left[ (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + (-1)^3 \binom{n}{3} + (-1)^4 \binom{n}{4} + \ldots + (-1)^n \binom{n}{n} + (1 - 1) \right]$$  \hspace{1cm} (31.41)$$

$$= (-1) \left[ (-1)^0 \binom{n}{0} + (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + (-1)^3 \binom{n}{3} + (-1)^4 \binom{n}{4} + \ldots + (-1)^n \binom{n}{n} - 1 \right]$$  \hspace{1cm} (31.42)$$

Thus, the contribution for any particular $x$ is one both on the l.h.s. and on the r.h.s. of the equation.
Inclusion/Exclusion

For the r.h.s. of Equation 31.37, we want to look what the contribution for this particular $x$ will be, and it is:

\[
n - \binom{n}{2} + \binom{n}{3} - \binom{n}{4} + \ldots + (-1)^{n-1} \binom{n}{n} = (-1)^0 \binom{n}{1} + (-1)^1 \binom{n}{2} + (-1)^2 \binom{n}{3} + (-1)^3 \binom{n}{4} + \ldots + (-1)^{n-1} \binom{n}{n}
\] (31.38)

\[
= (-1) \left[ (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + (-1)^3 \binom{n}{3} + (-1)^4 \binom{n}{4} + \ldots + (-1)^n \binom{n}{n} \right]
\] (31.39)

\[
= (-1) \left[ (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + (-1)^3 \binom{n}{3} + (-1)^4 \binom{n}{4} + \ldots + (-1)^n \binom{n}{n} + (1-1) \right]
\] (31.40)

\[
= (-1) \left[ (-1)^0 \binom{n}{0} + (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + (-1)^3 \binom{n}{3} + (-1)^4 \binom{n}{4} + \ldots + (-1)^n \binom{n}{n} - 1 \right]
\] (31.41)

\[
= (-1) \left[ (1-1)^n - 1 \right]
\] (31.42)

Thus, the contribution for any particular $x$ is one both on the l.h.s. and on the r.h.s. of the equation.
For the r.h.s. of Equation 31.37, we want to look what the contribution for this particular \( x \) will be, and it is:

\[
\begin{align*}
\text{n} & - \binom{n}{2} + \binom{n}{3} - \binom{n}{4} + \ldots + (-1)^{n-1} \binom{n}{n} \\
& = (-1)^0 \binom{n}{1} + (-1)^1 \binom{n}{2} + (-1)^2 \binom{n}{3} + (-1)^3 \binom{n}{4} + \ldots + (-1)^{n-1} \binom{n}{n} \\
& = (-1) \left[ (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + (-1)^3 \binom{n}{3} + (-1)^4 \binom{n}{4} + \ldots + (-1)^n \binom{n}{n} \right] \\
& = (-1) \left[ (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + (-1)^3 \binom{n}{3} + (-1)^4 \binom{n}{4} + \ldots + (-1)^n \binom{n}{n} + (1 - 1) \right] \\
& = (-1) \left[ (-1)^0 \binom{n}{0} + (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + (-1)^3 \binom{n}{3} + (-1)^4 \binom{n}{4} + \ldots + (-1)^n \binom{n}{n} - 1 \right] \\
& = (-1) \left[ (1 - 1)^n - 1 \right] = 1
\end{align*}
\]

Thus, the contribution for any particular \( x \) is one both on the l.h.s. and on the r.h.s. of the equation.
Next, suppose that $x \in A_i$ for $i \in S$, where $|S| = k$. In other words, $x$ is in only exactly $k$ of the sets $A_k$ rather than all of them, where $0 \leq k < n$. 

Reminder of the inclusion/exclusion formula.

$$|\bigcap_{i=1}^{n} A_i| = \sum_{i=1}^{n} |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cup A_j| + \sum_{1 \leq i < j < k \leq n} |A_i \cup A_j \cup A_k| - \ldots + (-1)^{n-1} |A_1 \cup A_2 \cup \ldots \cup A_n|$$ (31.46)

The l.h.s of Equation 31.46 now contributes only 0 (zero) for this $x$. For the r.h.s of Equation 31.46, then the contribution for this particular $x$ will be . . .
Next, suppose that \( x \in A_i \) for \( i \in S \), where \( |S| = k \). In other words, \( x \) is in only exactly \( k \) of the sets \( A_k \) rather than all of them, where \( 0 \leq k < n \).

Reminder of the inclusion/exclusion formula.

\[
| \cap_{i=1}^{n} A_i | = \sum_{i=1}^{n} |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cup A_j| + \sum_{1 \leq i < j < k \leq n} |A_i \cup A_j \cup A_k| + \ldots + (-1)^{n-1}|A_1 \cup A_2 \cup \ldots \cup A_n| \tag{31.44}
\]
Next, suppose that \( x \in A_i \) for \( i \in S \), where \( |S| = k \). In other words, \( x \) is in only exactly \( k \) of the sets \( A_k \) rather than all of them, where \( 0 \leq k < n \).

Reminder of the inclusion/exclusion formula.

\[
| \cap_{i=1}^{n} A_i | = \sum_{i=1}^{n} |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cup A_j| \\
+ \sum_{1 \leq i < j < k \leq n} |A_i \cup A_j \cup A_k| + \ldots \\
+ (-1)^{n-1}|A_1 \cup A_2 \cup \ldots \cup A_n| 
\]  

(31.44)  

(31.45)  

(31.46)

The l.h.s. of Equation 31.46 now contributes only 0 (zero) for this \( x \).
Next, suppose that \( x \in A_i \) for \( i \in S \), where \( |S| = k \). In other words, \( x \) is in only exactly \( k \) of the sets \( A_k \) rather than all of them, where \( 0 \leq k < n \).

Reminder of the inclusion/exclusion formula.

\[
|\cap_{i=1}^n A_i| = \sum_{i=1}^n |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cup A_j| + \sum_{1 \leq i < j < k \leq n} |A_i \cup A_j \cup A_k| + \ldots + (-1)^{n-1}|A_1 \cup A_2 \cup \ldots \cup A_n| \tag{31.44}
\]

The l.h.s. of Equation 31.46 now contributes only 0 (zero) for this \( x \).

For the r.h.s. of Equation 31.46, then the contribution for this particular \( x \) will be . . .
For the r.h.s. of Equation 31.37, we have the contribution for this particular \( x \) be:

\[
k - \left[ \binom{n}{2} - \binom{n-k}{2} \right] + \left[ \binom{n}{3} - \binom{n-k}{3} \right] + \ldots + (-1)^{n-k-1} \left[ \binom{n}{n-k} - \binom{n-k}{n-k} \right]
+ (-1)^{n-k} \left[ \binom{n}{n-k+1} \right] + \ldots + (-1)^{n-1} \left[ \binom{n}{n} \right]
\]

\[
= (-1)^0 \left[ \binom{n}{1} - \binom{n-k}{1} \right] - \left[ \binom{n}{2} - \binom{n-k}{2} \right] + \left[ \binom{n}{3} - \binom{n-k}{3} \right] + \ldots + (-1)^{n-k-1} \left[ \binom{n}{n-k} - \binom{n-k}{n-k} \right]
+ (-1)^{n-k} \left[ \binom{n}{n-k+1} \right] + \ldots + (-1)^{n-1} \left[ \binom{n}{n} \right]
\]

\[
= (-1) \left[ (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + \ldots + (-1)^n \binom{n}{n} + 1 - 1 \right]
- \left[ (-1)^0 \binom{n-k}{1} + (-1)^1 \binom{n-k}{2} + (-1)^2 \binom{n-k}{3} + \ldots + (-1)^{n-k-1} \binom{n-k}{n-k} \right]
\]

\[
= (-1) \left[ (-1)^0 \binom{n}{0} + (-1)^1 \binom{n}{1} + (-1)^2 \binom{n}{2} + \ldots + (-1)^n \binom{n}{n} - 1 \right]
+ \left[ (-1)^0 \binom{n-k}{0} + (-1)^1 \binom{n-k}{1} + (-1)^2 \binom{n-k}{2} + (-1)^3 \binom{n-k}{3} + \ldots + (-1)^{n-k} \binom{n-k}{n-k} - 1 \right]
\]

\[
= (-1) \left[ (1 - 1)^n - 1 \right] + [ (1 - 1)^{n-k} - 1 ] = 1 - 1 = 0
\]
Inclusion/Exclusion, two forms

- The same exact argument can be used to show inclusion/exclusion formula for the signed measure $\mu$, i.e.,

$$
\mu(\cap_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i) - \sum_{1 \leq i < j \leq n} \mu(A_i \cup A_j) \\
+ \sum_{1 \leq i < j < k \leq n} \mu(A_i \cup A_j \cup A_k) + \ldots \\
+ (-1)^{n-1} \mu(A_1 \cup A_2 \cup \ldots \cup A_n)
$$

(31.52)
Inclusion/Exclusion, two forms

- The same exact argument can be used to show inclusion/exclusion formula for the signed measure \( \mu \), i.e.,

\[
\mu(\cap_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i) - \sum_{1 \leq i < j \leq n} \mu(A_i \cup A_j)
\]

\[+ \sum_{1 \leq i < j < k \leq n} \mu(A_i \cup A_j \cup A_k) + \ldots \]  

\[+ (-1)^{n-1} \mu(A_1 \cup A_2 \cup \ldots \cup A_n) \]  

- A “dual” form of inclusion/exclusion has the form:

\[
\mu(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i) - \sum_{1 \leq i < j \leq n} \mu(A_i \cap A_j)
\]

\[+ \sum_{1 \leq i < j < k \leq n} \mu(A_i \cap A_j \cap A_k) + \ldots \]

\[+ (-1)^{n-1} \mu(A_1 \cap A_2 \cap \ldots \cap A_n) \]
Another (easier?, shorter) way of writing these is as:

\[ \mu(\cap_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k+1} \left( \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \mu(A_{i_1} \cup \cdots \cup A_{i_k}) \right) \] (31.58)

and

\[ \mu(\cup_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k+1} \left( \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \mu(A_{i_1} \cap \cdots \cap A_{i_k}) \right) \] (31.59)
What needs to be specified: getting atoms from sets

**Theorem 31.3.1**

Signed measure $\mu$ on $\mathcal{F}_n$ is fully specified by $\{\mu(B) : B \in \mathcal{B}\}$ which can be *any* set of real numbers.

- So before, we defined $\mu$ on all atoms and noted that this allowed us to compute $\mu$ everywhere else.
- Here, in the above theorem, we are defining $\mu$ only on elements of $\mathcal{B}$ and are again saying that this allows us to compute the values of $\mu$ everywhere else.
- we will see that this allows us to generate all standard mutual-information quantities, and some other less standard ones, using just entropy to fill out $\mu$ on $\mathcal{B}$.
Proof of Theorem 31.3.1

Note that $|A| = |B| = 2^n - 1 \triangleq k$
Proof of Theorem 31.3.1.

- Note that $|A| = |B| = 2^n - 1 \triangleq k$
- Define $\vec{a} = [\ldots \mu(A) \ldots]^\top$ for all $A \in \mathcal{A}$. length($\vec{a}$) = $k$
Proof of Theorem 31.3.1.

- Note that $|A| = |B| = 2^n - 1 \triangleq k$
- Define $\vec{a} = [\ldots \mu(A) \ldots]^\top$ for all $A \in \mathcal{A}$. $\text{length}(\vec{a}) = k$
- Define $\vec{b} = [\ldots \mu(B) \ldots]^\top$ for all $B \in \mathcal{B}$. $\text{length}(\vec{b}) = k$

...
Proof of Theorem 31.3.1

Note that $|\mathcal{A}| = |\mathcal{B}| = 2^n - 1 \triangleq k$.

Define $\vec{a} = [\ldots \mu(A) \ldots]^\top$ for all $A \in \mathcal{A}$. $\text{length}(\vec{a}) = k$.

Define $\vec{b} = [\ldots \mu(B) \ldots]^\top$ for all $B \in \mathcal{B}$. $\text{length}(\vec{b}) = k$.

For any $B \in \mathcal{B}$, we have $B = \bigcup_{\ell \in \mathcal{A}(B)} A_\ell$ with $A_\ell \in \mathcal{A}$ and where $\mathcal{A}(B)$ are the indices of the atoms that comprise $B$. 

...
proof of Theorem 31.3.1.

- Note that $|A| = |B| = 2^n - 1 \triangleq k$
- Define $\vec{a} = [\ldots \mu(A) \ldots]^\top$ for all $A \in \mathcal{A}$. length($\vec{a}$) = k
- Define $\vec{b} = [\ldots \mu(B) \ldots]^\top$ for all $B \in \mathcal{B}$. length($\vec{b}$) = k
- For any $B \in \mathcal{B}$, we have $B = \bigcup_{\ell \in \mathcal{A}(B)} A_\ell$ with $A_\ell \in \mathcal{A}$ and where $\mathcal{A}(B)$ are the indices of the atoms that comprise $B$.
- Therefore, there exists a unique $k \times k$ matrix $C_n$ such that $\vec{b} = C_n \vec{a}$
Proof of Theorem 31.3.1

proof of Theorem 31.3.1.

- Note that $|A| = |B| = 2^n - 1 \triangleq k$
- Define $\vec{a} = [\ldots \mu(A) \ldots]^\top$ for all $A \in \mathcal{A}$. $\text{length}(\vec{a}) = k$
- Define $\vec{b} = [\ldots \mu(B) \ldots]^\top$ for all $B \in \mathcal{B}$. $\text{length}(\vec{b}) = k$
- For any $B \in \mathcal{B}$, we have $B = \bigcup_{\ell \in \mathcal{A}(B)} A_\ell$ with $A_\ell \in \mathcal{A}$ and where $\mathcal{A}(B)$ are the indices of the atoms that comprise $B$.
- Therefore, there exists a unique $k \times k$ matrix $C_n$ such that $\vec{b} = C_n \vec{a}$
- But also, we claim that any $\mu(A)$ for $A \in \mathcal{A}$ can be expressed as a linear combination of $\{\mu(B)\}_{B \in \mathcal{B}(A)}$ for the appropriate $\mathcal{B}(A) \subseteq \mathcal{B}$, and this can be done using inclusion/exclusion.
proof of Theorem 31.3.1.

Here, the inclusion/exclusion principle takes the form conditioned on (or excluding) $B$:

$$\mu(\cap_{k=1}^{n} \tilde{X}_k \setminus B) = \sum_{1 \leq i \leq n} \mu(\tilde{X}_i \setminus B) - \sum_{1 \leq i < j \leq n} \mu((\tilde{X}_i \cup \tilde{X}_j) \setminus B) + \cdots + (-1)^{n+1}\mu((\tilde{X}_1 \cup \tilde{X}_2 \cup \cdots \cup \tilde{X}_n) \setminus B)$$

(31.60)
Proof of Theorem 31.3.1

Here, the inclusion/exclusion principle takes the form conditioned on (or excluding) $B$:

$$\mu(\cap_{k=1}^{n}\tilde{X}_k \setminus B) = \sum_{1 \leq i \leq n} \mu(\tilde{X}_i \setminus B) - \sum_{1 \leq i < j \leq n} \mu((\tilde{X}_i \cup \tilde{X}_j) \setminus B)$$

$$+ \cdots + (-1)^{n+1} \mu((\tilde{X}_1 \cup \tilde{X}_2 \cup \cdots \cup \tilde{X}_n) \setminus B)$$

(31.60)

How can this help us? Note: this formula works for any number of sets, not just all $n$ of them (so we can take $n$ to be any number of sets rather than only all of them). 

...
Proof of Theorem 31.3.1

Since $A \setminus B = A \cap B^c$, every atom $A \in \mathcal{A}$ corresponds to:

$$A = \bigcap_{i=1}^{n} Y_i = \left( \bigcap_{j:Y_j=\tilde{X}_j} \tilde{X}_j \right) \cap \left( \bigcap_{j:Y_j=\tilde{X}_j^c} \tilde{X}_j^c \right)$$

$$= \left( \bigcap_{j:Y_j=\tilde{X}_j} \tilde{X}_j \right) \cap \left( \bigcup_{j:Y_j=\tilde{X}_j} \tilde{X}_j \right)^c$$

$$= \left( \bigcap_{j:Y_j=\tilde{X}_j} \tilde{X}_j \right) \setminus \left( \bigcup_{j:Y_j=\tilde{X}_j^c} \tilde{X}_j \right) = \left( \bigcap_{j:Y_j=\tilde{X}_j} \tilde{X}_j \right) \setminus B$$

...
proof of Theorem 31.3.1.

- Also, each of the terms of the r.h.s. of the inclusion/exclusion formula (Eqn.(31.60)) may take the form:

\[ \mu(\tilde{X}_i \cup \tilde{X}_j \cup \cdots \cup \tilde{X}_k \setminus B) = \mu(\tilde{X}_i \cup \tilde{X}_j \cup \cdots \cup \tilde{X}_k \setminus \bigcup_{\ell} \tilde{X}_\ell) \]

\[ = \mu(\tilde{X}_i \cup \tilde{X}_j \cup \cdots \cup \tilde{X}_k \cup \bigcup_{\ell} \tilde{X}_\ell) - \mu(\bigcup_{\ell} \tilde{X}_\ell) \]  

(31.64)

(31.65)

which is true since \( \mu(U \setminus V) = \mu(U \cup V) - \mu(V), \forall U, V. \)
Proof of Theorem 31.3.1

Also, each of the terms of the r.h.s. of the inclusion/exclusion formula (Eqn.(31.60)) may take the form:

\[
\mu(\tilde{X}_i \cup \tilde{X}_j \cup \cdots \cup \tilde{X}_k \setminus B) = \mu(\tilde{X}_i \cup \tilde{X}_j \cup \cdots \cup \tilde{X}_k \setminus \bigcup_{\ell} \tilde{X}_\ell)
\]

(31.64)

\[
= \mu(\tilde{X}_i \cup \tilde{X}_j \cup \cdots \cup \tilde{X}_k \cup \bigcup_{\ell} \tilde{X}_\ell) - \mu(\bigcup_{\ell} \tilde{X}_\ell)
\]

(31.65)

which is true since \(\mu(U \setminus V) = \mu(U \cup V) - \mu(V), \forall U, V\).

Thus, the measure of any atom \(A \in \mathcal{A}\) is representable as a sum of weighted measures of the unions of the basic sets \(\mathcal{B}\)!
Proof of Theorem 31.3.1

Also, each of the terms of the r.h.s. of the inclusion/exclusion formula (Eqn.(31.60)) may take the form:

\[
\mu(\tilde{X}_i \cup \tilde{X}_j \cup \cdots \cup \tilde{X}_k \setminus B) = \mu(\tilde{X}_i \cup \tilde{X}_j \cup \cdots \cup \tilde{X}_k \setminus \bigcup_{\ell} \tilde{X}_\ell)
\]

(31.64)

\[
= \mu(\tilde{X}_i \cup \tilde{X}_j \cup \cdots \cup \tilde{X}_k \cup \bigcup_{\ell} \tilde{X}_\ell) - \mu(\bigcup_{\ell} \tilde{X}_\ell)
\]

(31.65)

which is true since \(\mu(U \setminus V) = \mu(U \cup V) - \mu(V), \ \forall U, V.\)

Thus, the measure of any atom \(A \in \mathcal{A}\) is representable as a sum of weighted measures of the unions of the basic sets \(B!\)

Therefore, there exists a \(k \times k\) matrix \(D_n\) such that \(\vec{a} = D_n \vec{b}\)

(before we had \(\vec{b} = C_n \vec{a}\).)

\[
\ldots
\]
proof of Theorem 31.3.1.

Since $C_n$ is unique, so is $D_n$ with $D_n = C^{-1}_n$. 
Proof of Theorem 31.3.1

Since $C_n$ is unique, so is $D_n$ with $D_n = C_n^{-1}$.

To summarize, we can define measure values only on $\mathcal{B}$ and it defines the measures for all elements of $\mathcal{F}_n$. 
proof of Theorem 31.3.1.

- Since $C_n$ is unique, so is $D_n$ with $D_n = C_n^{-1}$.

- To summarize, we can define measure values only on $\mathcal{B}$ and it defines the measures for all elements of $\mathcal{F}_n$.

- For example, we can define just the values $H(X_G)$ for $G \subseteq [n]$ and this defines every other information theoretic value.
Examples

- Here are two examples of the above ideas, expressible as lemmas, first proven using set theory and second proven using information theoretic ideas.
Examples

Here are two examples of the above ideas, expressible as lemmas, first proven using set theory and second proven using information theoretic ideas.

Let $A, B, C$ be sets.
Examples

- Here are two examples of the above ideas, expressible as lemmas, first proven using set theory and second proven using information theoretic ideas.

- Let $A, B, C$ be sets.

- **Lemma:**
  
  $$\mu((A \cap B) \setminus C) = \mu(A \cup C) + \mu(B \cup C) - \mu(A \cup B \cup C) - \mu(C').$$
Here are two examples of the above ideas, expressible as lemmas, first proven using set theory and second proven using information theoretic ideas.

Let $A, B, C$ be sets.

**lemma:**
\[
\mu((A \cap B) \setminus C) = \mu(A \cup C) + \mu(B \cup C) - \mu(A \cup B \cup C) - \mu(C').
\]

**proof:**
\[
\begin{align*}
\mu((A \cap B) \setminus C) &= \mu(A \setminus C) + \mu(B \setminus C) - \mu((A \cup B) \setminus C) \\
&= (\mu(A \cup C) - \mu(C')) + (\mu(B \cup C) - \mu(C')) \\
&\quad - (\mu(A \cup B \cup C) - \mu(C')) \\
&= \mu(A \cup C) + \mu(B \cup C) - \mu(A \cup B \cup C) - \mu(C)
\end{align*}
\]
As a Shannon measure (let $A, B, C$ be random variables):

\[
I(A; B | C) = H(A, C) + H(B, C) - H(A, B, C) - H(C)
\]

\[
= H(A, C) + H(B, C) - H(A, B, C) + H(C)
\]

Key point: this is the same as before, but in the first case we showed it using set theory.
As a Shannon measure (let $A, B, C$ be random variables):

**Lemma:** $I(A; B|C) = H(A, C) + H(B, C) - H(A, B, C) - H(C)$

Key point: this is the same as before, but in the first case we showed it using set theory.
As a Shannon measure (let $A, B, C$ be random variables):

**lemma:** $I(A; B|C) = H(A, C) + H(B, C) - H(A, B, C) - H(C)$

**proof:**

$$I(A; B|C) = H(A|C) - H(A|B, C)$$  \hspace{1cm} (31.69)

$$= H(A, C) - H(C) - (H(A, B, C) - H(B, C))$$  \hspace{1cm} (31.70)

$$= H(A, C) + H(B, C) - H(A, B, C) + H(C)$$  \hspace{1cm} (31.71)
Examples

- As a Shannon measure (let $A, B, C$ be random variables):
  - lemma: $I(A; B|C) = H(A, C) + H(B, C) - H(A, B, C) - H(C)$
  - proof:

$$I(A; B|C) = H(A|C) - H(A|B, C)$$

$$= H(A, C) - H(C) - (H(A, B, C) - H(B, C))$$

$$= H(A, C) + H(B, C) - H(A, B, C) + H(C)$$

- Key point: this is the same as before, but in the first case we showed it using set theory.
Another key point. The information measure generates everything. That is, assigning the measures only for $G$ as in

$$\mu(\tilde{X}_G) \triangleq H(X_G) \text{ for } G \subseteq [n]$$

(31.72)

is sufficient to get all remaining Shannon’s information theoretic quantities.
Another key point. The information measure generates everything. That is, assigning the measures only for $G$ as in

$$\mu(\tilde{X}_G) \triangleq H(X_G) \text{ for } G \subseteq [n]$$

(31.72)

is sufficient to get all remaining Shannon’s information theoretic quantities.

Example: Start by defining $\mu(\tilde{X}_G)$ only on $B$ with entropic quantities as above.
Another key point. The information measure generates everything. That is, assigning the measures only for $G$ as in

$$
\mu(\tilde{X}_G) \triangleq H(X_G) \text{ for } G \subseteq [n]
$$

(31.72)

is sufficient to get all remaining Shannon’s information theoretic quantities.

Example: Start by defining $\mu(\tilde{X}_G)$ only on $B$ with entropic quantities as above.

Then consider $G$, $G'$, and $G''$ as sets, and form

$$
\mu((\tilde{X}_G \cap \tilde{X}_{G'}) \setminus \tilde{X}_{G''})
= \mu(\tilde{X}_G \cup \tilde{X}_{G''}) + \mu(\tilde{X}_{G'} \cup \tilde{X}_{G''}) - \mu(X_G, X_{G'}, X_{G''}) - \mu(X_{G''})
= I(X_G; X_{G'}|X_{G''})
$$

(31.74)
Another key point. The information measure generates everything. That is, assigning the measures only for $G$ as in

$$\mu(\tilde{X}_G) \triangleq H(X_G) \text{ for } G \subseteq [n]$$  \hspace{1cm} (31.72)

is sufficient to get all remaining Shannon’s information theoretic quantities.

Example: Start by defining $\mu(\tilde{X}_G)$ only on $\mathcal{B}$ with entropic quantities as above.

Then consider $G$, $G'$, and $G''$ as sets, and form

$$\mu((\tilde{X}_G \cap \tilde{X}_{G'}) \setminus \tilde{X}_{G''})$$  \hspace{1cm} (31.73)

$$= \mu(\tilde{X}_G \cup \tilde{X}_{G''}) + \mu(\tilde{X}_{G'} \cup \tilde{X}_{G''}) - \mu(X_G, X_{G'}, X_{G''}) - \mu(X_{G''})$$

$$= I(X_G; X_{G'}|X_{G''})$$  \hspace{1cm} (31.74)

That is, in most general case, we get conditional mutual information, but setting various combinations of $G$, $G'$, or $G''$ to empty allows us to get $I(X_G; X_{G''})$ (if $G'' = \emptyset$), $H(X_G|X_{G''})$ (if $G' = \emptyset$), etc.
Signed Measure

- When \( n = 2 \) all measures are positive given entropic instantiations (e.g., discrete entropy and MI are non-negative).
Signed Measure

- When $n = 2$ all measures are positive given entropic instantiations (e.g., discrete entropy and MI are non-negative).
- When $n = 3$, is this so?
Signed Measure

- When \( n = 2 \) all measures are positive given entropic instantiations (e.g., discrete entropy and MI are non-negative).
- When \( n = 3 \), is this so? What is:

\[
\mu(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3)
\]  

(31.75)

and what might it mean?
Signed Measure

- When \( n = 2 \) all measures are positive given entropic instantiations (e.g., discrete entropy and MI are non-negative).
- When \( n = 3 \), is this so? What is:

\[
\mu(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) \quad (31.75)
\]

and what might it mean?
- We have that:

\[
\mu(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) + \mu(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3^c) = \mu(\tilde{X}_1 \cap \tilde{X}_2) \quad (31.76)
\]
Signed Measure

- When $n = 2$ all measures are positive given entropic instantiations (e.g., discrete entropy and MI are non-negative).

- When $n = 3$, is this so? What is:

  $$\mu(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3)$$  \hspace{1cm} (31.75)

  and what might it mean?

- We have that:

  $$\mu(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) + \mu(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3^c) = \mu(\tilde{X}_1 \cap \tilde{X}_2)$$  \hspace{1cm} (31.76)

  which translates as

  $$I(X_1; X_2; X_3) + I(X_1; X_2|X_3) = I(X_1; X_2)$$  \hspace{1cm} (31.77)

  defined quantity
In fact, we have that:

\[ I(X_1; X_2; X_3) = I(X_1; X_2) - I(X_1; X_2|X_3) \]  
(31.78)

\[ = I(X_1; X_3) - I(X_1; X_3|X_2) \]  
(31.79)

\[ = I(X_2; X_3) - I(X_2; X_3|X_1) \]  
(31.80)

Why is this?
In fact, we have that:

\[
I(X_1; X_2; X_3) = I(X_1; X_2) - I(X_1; X_2|X_3) \tag{31.78}
\]

\[
= I(X_1; X_3) - I(X_1; X_3|X_2) \tag{31.79}
\]

\[
= I(X_2; X_3) - I(X_2; X_3|X_1) \tag{31.80}
\]

Why is this? Due to the three ways of “de-eliminating” or “de-marginalizing” the three variables, that is:

\[
\mu(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) + \mu(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3^c) = \mu(\tilde{X}_1 \cap \tilde{X}_2) \tag{31.81}
\]

\[
\mu(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) + \mu(\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3) = \mu(\tilde{X}_1 \cap \tilde{X}_3) \tag{31.82}
\]

\[
\mu(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) + \mu(\tilde{X}_1^c \cap \tilde{X}_2 \cap \tilde{X}_3) = \mu(\tilde{X}_2 \cap \tilde{X}_3) \tag{31.83}
\]