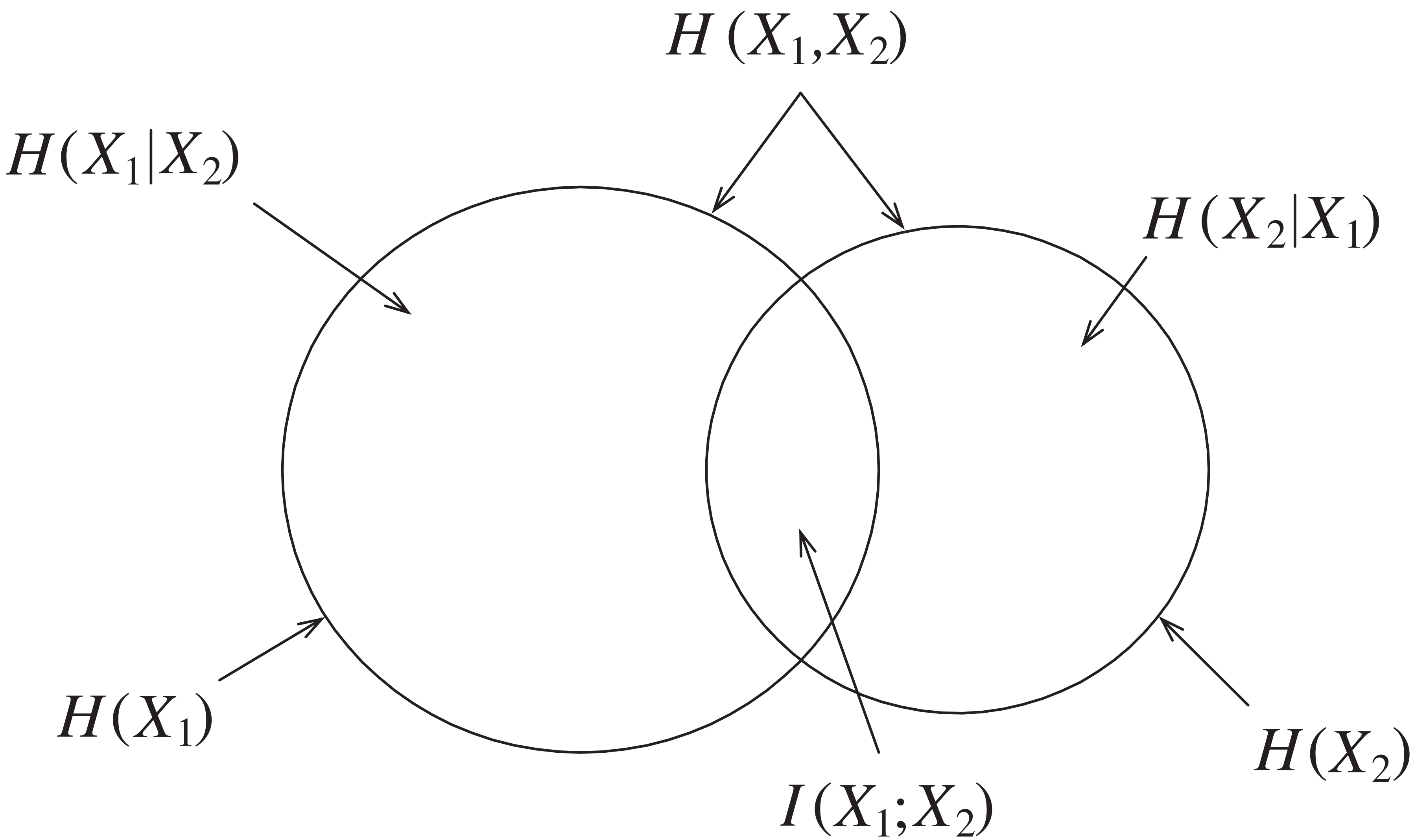
Chapter 3
The *I*-Measure

Department of Information Engineering
The Chinese University of Hong Kong

An Example



Substitution of Symbols

$$\begin{array}{ccc} H/I & \leftrightarrow & \mu^* \\ , & \leftrightarrow & \cup \\ ; & \leftrightarrow & \cap \\ | & \leftrightarrow & - \end{array}$$

- μ^* is some signed measure on $\mathcal{F}_n$.
- Examples:

  1.

$$\begin{aligned} \mu^*(\tilde{X}_1 - \tilde{X}_2) &= H(X_1|X_2) \\ \mu^*(\tilde{X}_2 - \tilde{X}_1) &= H(X_2|X_1), \\ \mu^*(\tilde{X}_1 \cap \tilde{X}_2) &= I(X_1;X_2) \end{aligned}$$

2. Inclusion-Exclusion formulation in set-theory

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

corresponds to

$$H(X_1, X_2) = H(X_1) + H(X_2) - I(X_1; X_2)$$

in information theory.

3.1 Preliminaries

Definition 3.1 The field $\mathcal{F}_n$ generated by sets $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_n$ is the collection of sets which can be obtained by any sequence of usual set operations (union, intersection, complement, and difference) on $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_n$.

Definition 3.2 The atoms of $\mathcal{F}_n$ are sets of the form $\cap_{i=1}^n Y_i$, where Y_i is either $\tilde{X}_i$ or $\tilde{X}_i^c$, the complement of $\tilde{X}_i$.

Example 3.3

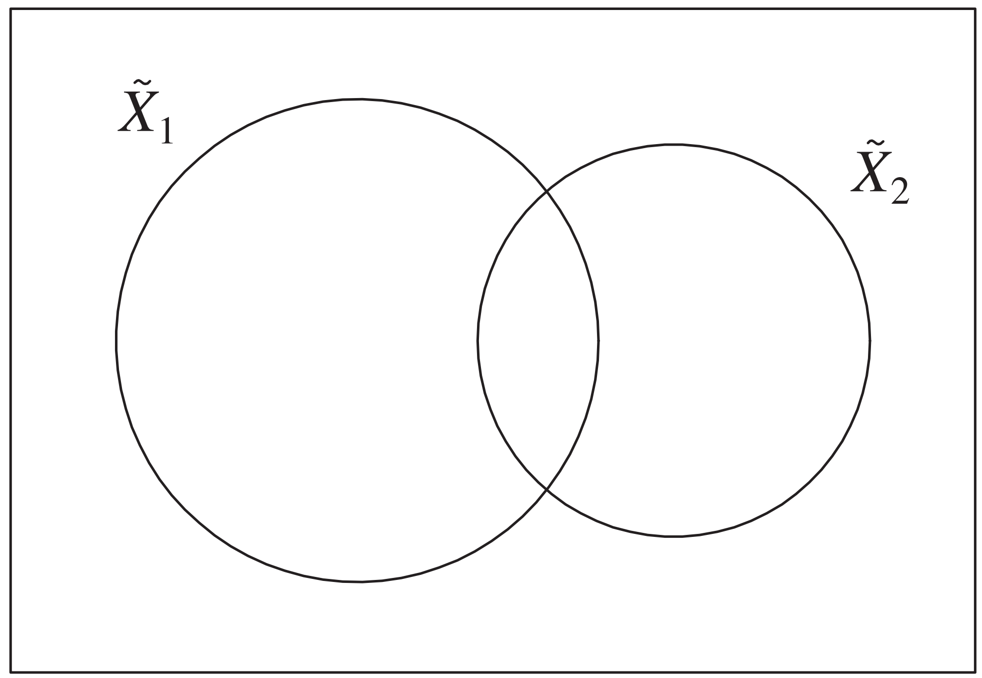
- The sets $\tilde{X}_1$ and $\tilde{X}_2$ generate the field $\mathcal{F}_2$.
- There are 4 atoms in $\mathcal{F}_2$.
- There are a total of 16 sets in $\mathcal{F}_2$

Definition 3.4 A real function μ defined on $\mathcal{F}_n$ is called a signed measure if it is set-additive, i.e., for disjoint A and B in $\mathcal{F}_n$,

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

Remark $\mu(\emptyset) = 0$.

Example 3.5



- A signed measure μ on $\mathcal{F}_2$ is completely specified by the values on the atoms
$$\mu(\tilde{X}_1 \cap \tilde{X}_2),\ \mu(\tilde{X}_1^c \cap \tilde{X}_2),\ \mu(\tilde{X}_1 \cap \tilde{X}_2^c),\ \mu(\tilde{X}_1^c \cap \tilde{X}_2^c)$$

- The value of μ on other sets in $\mathcal{F}_2$ are obtained by set-additivity.

Section 3.3
Construction of the *I*-Measure μ*

- Let $\tilde{X}$ be a set corresponding to a r.v. X.
- $\mathcal{N}_n = \{1, 2, \cdots, n\}$.
- Universal set
$$\Omega = \bigcup_{i \in \mathcal{N}_n} \tilde{X}_i.$$
- Empty atom of $\mathcal{F}_n$
$$A_0 = \bigcap_{i \in \mathcal{N}_n} \tilde{X}_i^c$$
- $\mathcal{A}$ is the set of other atoms of $\mathcal{F}_n$, called non-empty atoms. $|\mathcal{A}| = 2^n - 1$.
- A signed measure μ on $\mathcal{F}_n$ is completely specified by the values of μ on the nonempty atoms of $\mathcal{F}_n$.

Notations For nonempty subset G of $\mathcal{N}_n$:

- $X_G = (X_i, i \in G)$
- $\tilde{X}_G = \cup_{i \in G} \tilde{X}_i$

Theorem 3.6 Let

$$\mathcal{B} = \left\{ \tilde{X}_G : \ G \text{ is a nonempty subset of } \mathcal{N}_n \right\}.$$

Then a signed measure μ on $\mathcal{F}_n$ is completely specified by $\{\mu(B), B \in \mathcal{B}\}$, which can be any set of real numbers.

Proof of Theorem 3.6

- $|\mathcal{A}| = |\mathcal{B}| = 2^n - 1$
- $\mathbf{u}$ – column k-vector of $\mu(A), A \in \mathcal{A}$
- $\mathbf{h}$ – column k-vector of $\mu(B), B \in \mathcal{B}$
- Obviously can write $\mathbf{h} = C_n\mathbf{u}$, where C_n is a *unique* $k \times k$ matrix.
- On the other hand, for each $A \in \mathcal{A}$, $\mu(A)$ can be expressed as a linear combination of $\mu(B), B \in \mathcal{B}$ by applying

$$\begin{aligned} \mu(A \cap B - C) &= \mu(A - C) + \mu(B - C) - \mu(A \cup B - C) \\ \mu(A - B) &= \mu(A \cup B) - \mu(B). \end{aligned}$$

  (see Appendix 3.A) That is, $\mathbf{u} = D_n\mathbf{h}$.

- Then $\mathbf{u} = (D_n C_n)\mathbf{u}$, showing that $D_n = (C_n)^{-1}$ is unique.

Two Lemmas

Lemma 3.7

$$\mu(A \cap B - C) = \mu(A \cup C) + \mu(B \cup C) - \mu(A \cup B \cup C) - \mu(C).$$

Lemma 3.8

$$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z).$$

- Construct the I-Measure μ^* on $\mathcal{F}_n$ using by defining

$$\mu^*(\tilde{X}_G) = H(X_G)$$

for all nonempty subsets G of $\mathcal{N}_n$.

- μ^* is meaningful if it is consistent with all Shannon's information measures via the substitution of symbols, i.e., the following must hold for all (not necessarily disjoint) subsets G, G', G'' of $\mathcal{N}_n$ where G and G' are nonempty:

$$\mu^*(\tilde{X}_G \cap \tilde{X}_{G'} - \tilde{X}_{G''}) = I(X_G; X_{G'}|X_{G''})$$

- $\underline{G'' = \emptyset}$

$$\mu^*(\tilde{X}_G \cap \tilde{X}_{G'}) = I(X_G; X_{G'})$$

$\underline{G = G'}$

$$\mu^*(\tilde{X}_G - \tilde{X}_{G''}) = H(X_G|X_{G''})$$

$\underline{G = G' \text{ and } G'' = \emptyset}$

$$\mu^*(\tilde{X}_G) = H(X_G)$$

Theorem 3.9 μ^* is the unique signed measure on $\mathcal{F}_n$ which is consistent with all Shannon's information measures.

Implications

- Can formally regard Shannon's information measures for n r.v.'s as the unique signed measure μ^* defined on $\mathcal{F}_n$.
- Can employ set-theoretic tools to manipulate expressions of Shannon's information measures.

Proof of Theorem 3.9

- $$\begin{aligned}
&\mu^*(\tilde{X}_G \cap \tilde{X}_{G'} - \tilde{X}_{G''}) \\
&\quad = \mu^*(\tilde{X}_{G\cup G''}) + \mu^*(\tilde{X}_{G'\cup G''}) - \mu^*(\tilde{X}_{G\cup G'\cup G''}) - \mu^*(\tilde{X}_{G''}) \\
&\quad = H(X_{G\cup G''}) + H(X_{G'\cup G''}) - H(X_{G\cup G'\cup G''}) - H(X_{G''}) \\
&\quad = I(X_G; X_{G'}|X_{G''}),
\end{aligned}$$

- In order that μ^* is consistent with all Shannon's information measures,

$$\mu^*(\tilde{X}_G) = H(X_G)$$

  for all nonempty subsets G of $\mathcal{N}_n$.

- Thus μ^* is the unique signed measure on $\mathcal{F}_n$ which is consistent with all Shannon's information measures.

3.4 μ* can be Negative

- μ^* is nonnegative for $n = 2$.
- For $n = 3$, $\mu^*(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) = I(X_1; X_2; X_3)$ can be negative.

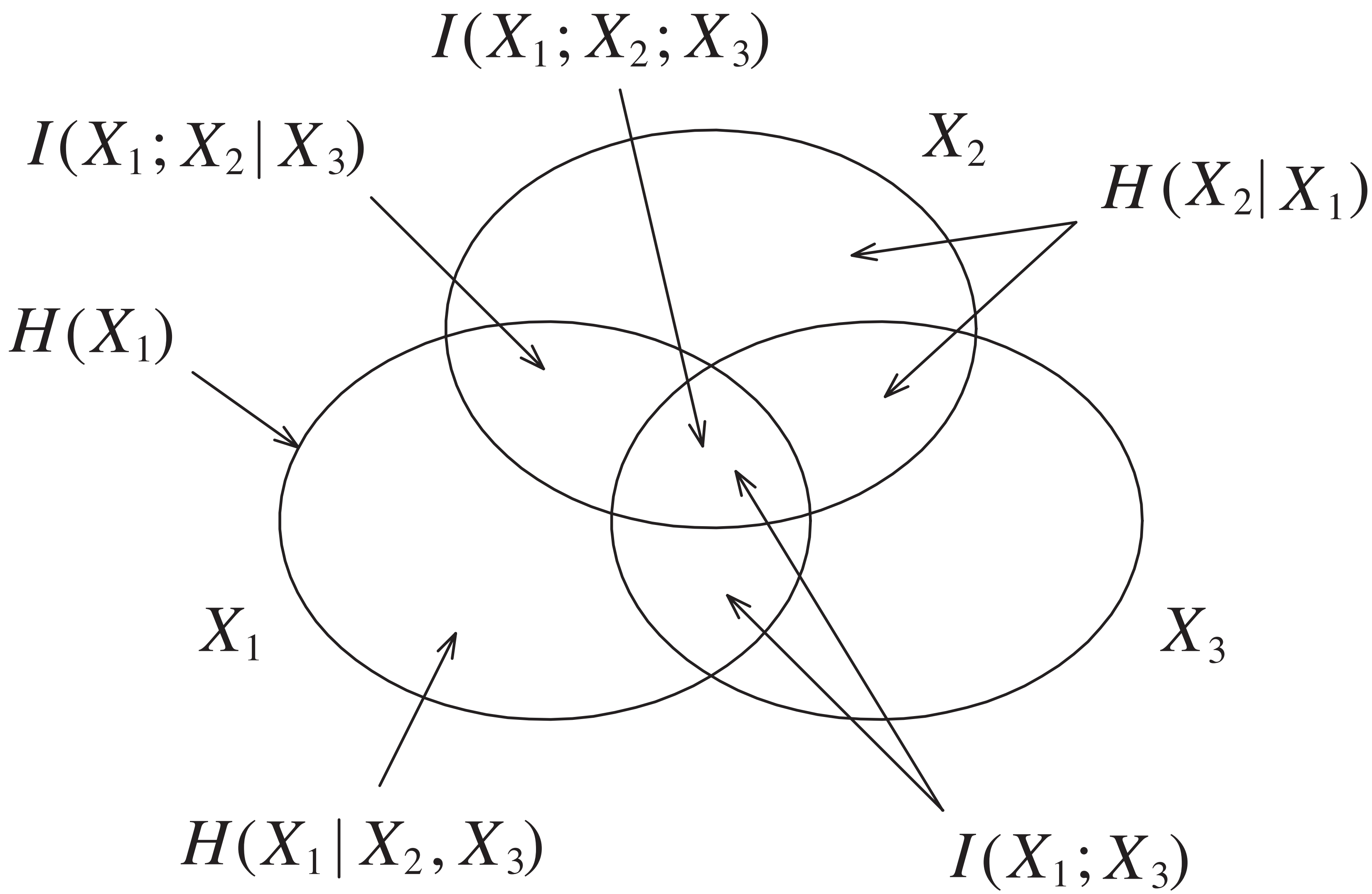
Example 3.10

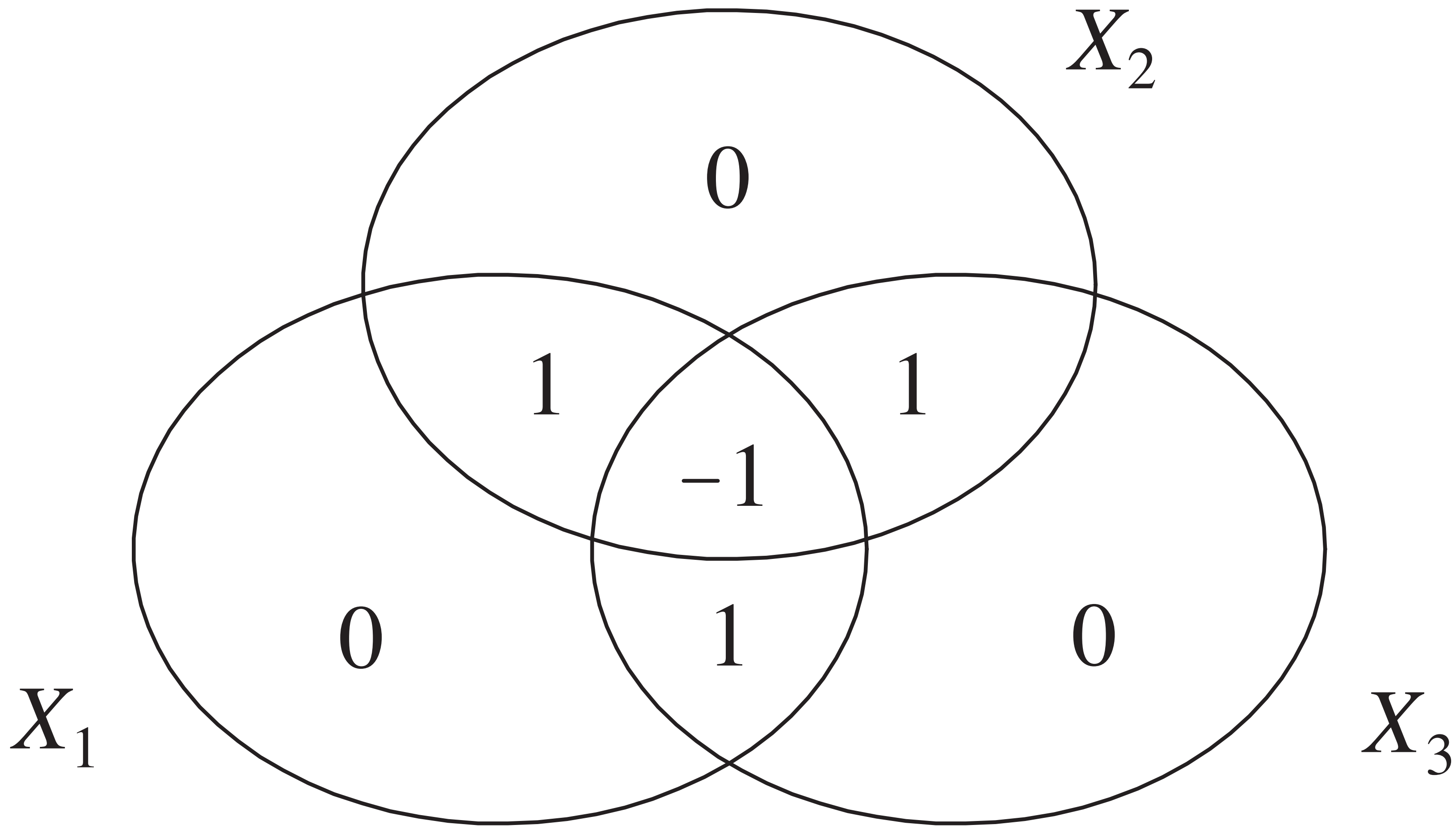
- X_1, X_2 – i.i.d. binary r.v.'s uniform on $\{0, 1\}$
- $X_3 = X_1 + X_2 \mod 2$
- Easy to check:
  - $H(X_i) = 1$, for all i
  - X_1, X_2, X_3 are pairwise independent, so that

$$H(X_i, X_j) = 2 \text{ and } I(X_i; X_j) = 0, \text{ for all } i \neq j$$

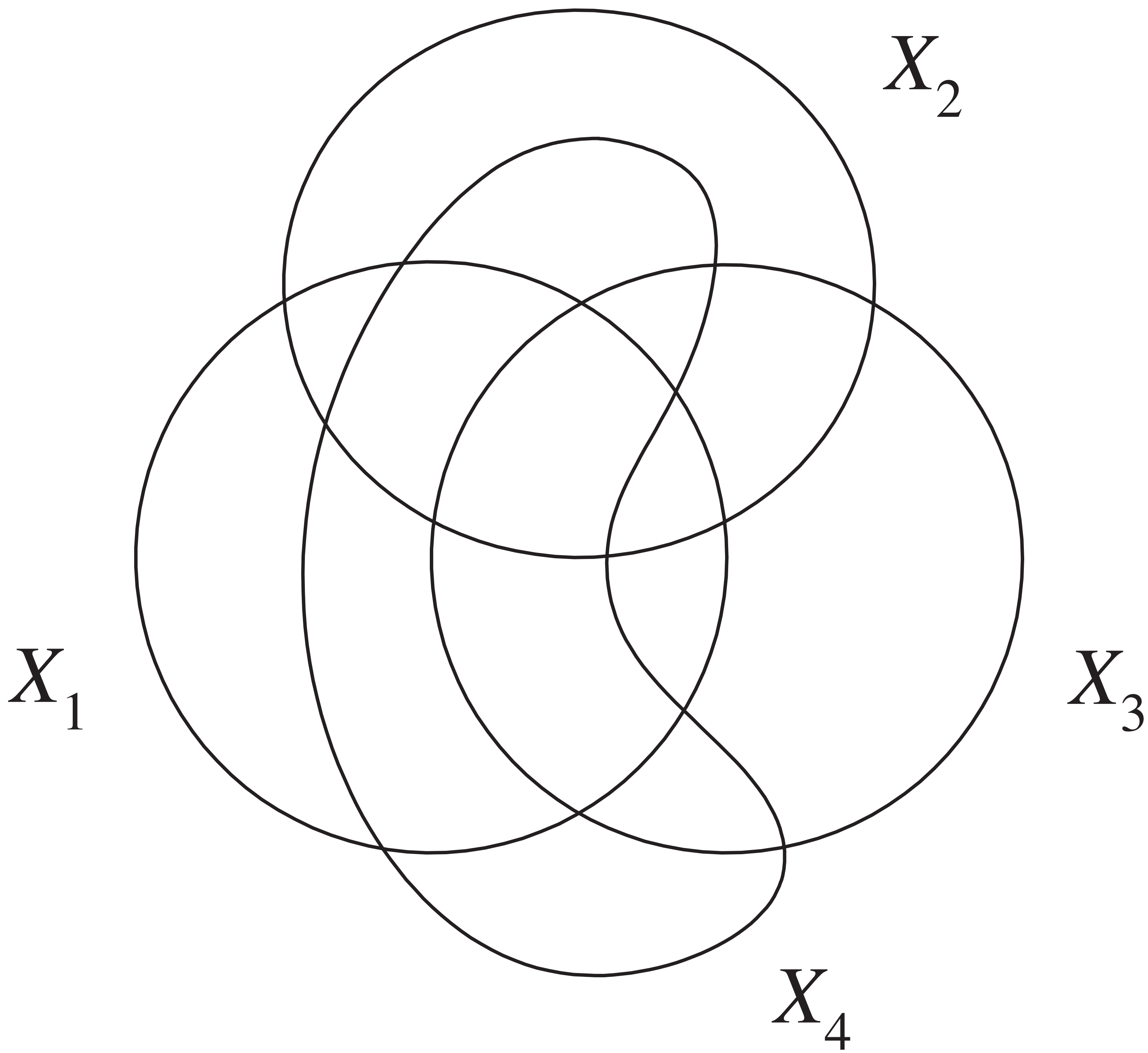
  - Under these constraints, $I(X_1; X_2; X_3) = -1$.

3.5 Information Diagrams





The information diagram for Example 3.10


X_2
X_1
X_3
X_4

Theorem 3.11 If there is no constraint on $X_1, X_2, \cdots, X_n$, then μ^* can take any set of nonnegative values on the nonempty atoms of $\mathcal{F}_n$.

Proof

- Let $Y_A, A \in \mathcal{A}$ be mutually independent r.v.'s.
- Define $X_i, i = 1, 2, \cdots, n$ by

$$X_i = (Y_A : A \in \mathcal{A} \text{ and } A \subset \tilde{X}_i).$$

- Claim: $X_1, X_2, \cdots, X_n$ so constructed induce the I-Measure μ^* such that

$$\mu^*(A) = H(Y_A), \text{ for all } A \in \mathcal{A}.$$

which are arbitrary nonnegative numbers.

- Consider

$$\begin{aligned} H(X_G) &= H(X_i, i \in G) \\ &= H((Y_A : A \in \mathcal{A} \text{ and } A \subset \tilde{X}_i), i \in G) \\ &= H(Y_A : A \in \mathcal{A} \text{ and } A \subset \tilde{X}_G) \\ &= \sum_{A \in \mathcal{A} : A \subset \tilde{X}_G} H(Y_A) \end{aligned}$$

- On the other hand,

$$H(X_G) = \mu^*(\tilde{X}_G) = \sum_{A \in \mathcal{A} : A \subset \tilde{X}_G} \mu^*(A)$$

- Thus

$$\sum_{A \in \mathcal{A} : A \subset \tilde{X}_G} H(Y_A) = \sum_{A \in \mathcal{A} : A \subset \tilde{X}_G} \mu^*(A)$$

- One solution is

$$\mu^*(A) = H(Y_A), \text{ for all } A \in \mathcal{A}.$$

- By the uniqueness of μ^*, this is the only solution.

Information Diagrams for Markov Chains

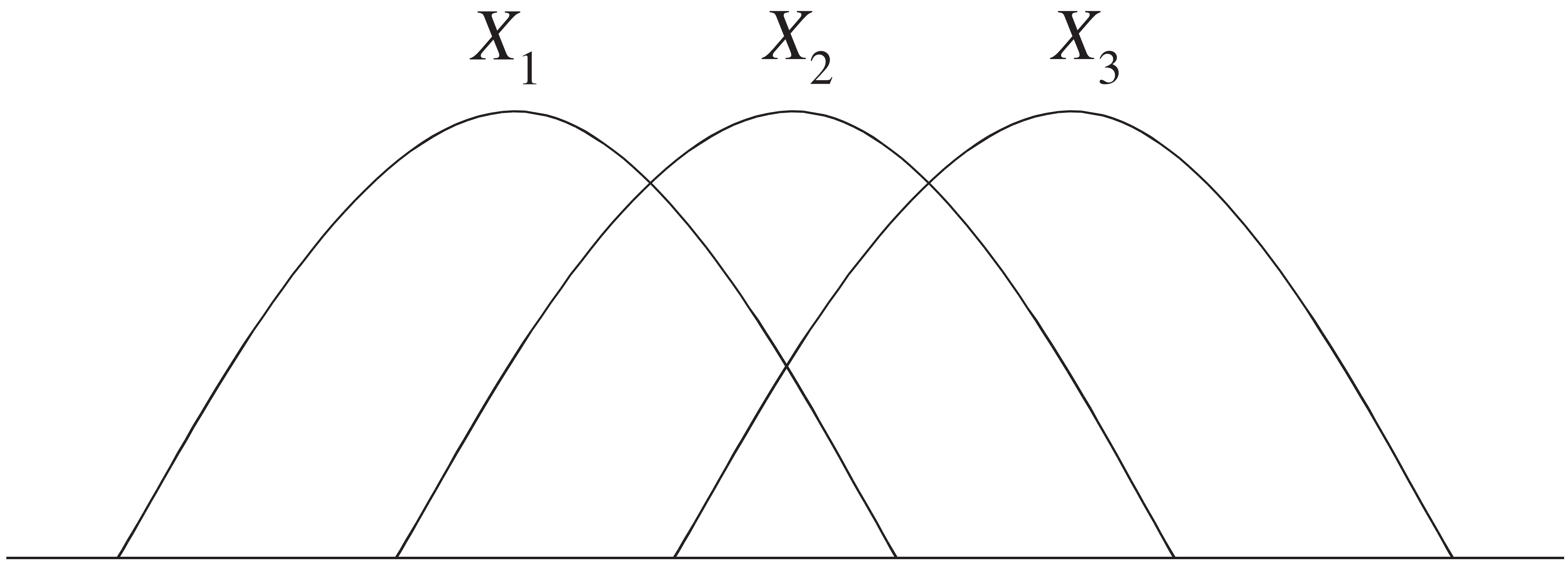
- If $X_1 \to X_2 \to \cdots \to X_n$ form a Markov chain, then the structure of μ^* is much simpler and hence the information diagram can be simplified.

- For $n = 3$, $X_1 \to X_2 \to X_3$ iff $I(X_1; X_3|X_2) = 0$. So the atom $\tilde{X}_1 \cap \tilde{X}_3 - \tilde{X}_2$ can be suppressed.

- The values of μ^* on the remaining atoms correspond to Shannon's information measures and hence are nonnegative. In particular,

$$\mu^*(\tilde{X}_1; \tilde{X}_2; \tilde{X}_3) = \mu^*(\tilde{X}_1; \tilde{X}_3) = I(X_1; X_3)$$

- Thus, μ^* is a measure.

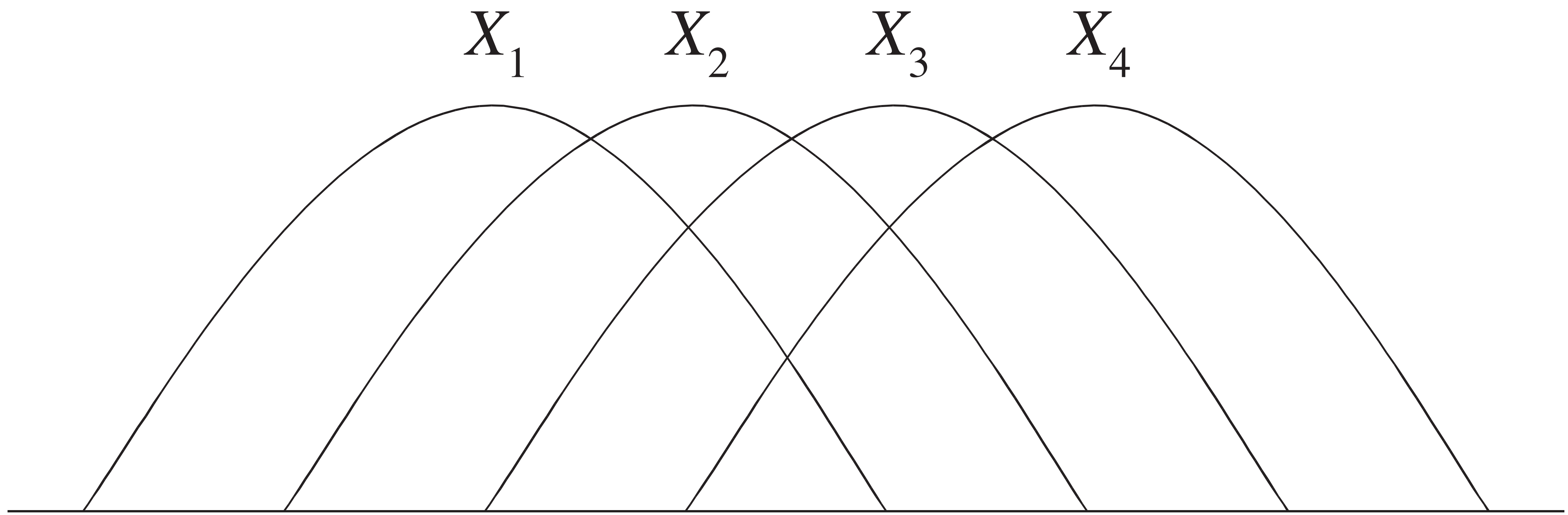

X_1
X_2
X_3

- For $n = 4$, μ^* vanishes on the following atoms:

$$\begin{array}{c}
\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3 \cap \tilde{X}_4^c \\
\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3 \cap \tilde{X}_4 \\
\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3^c \cap \tilde{X}_4 \\
\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3^c \cap \tilde{X}_4 \\
\tilde{X}_1^c \cap \tilde{X}_2 \cap \tilde{X}_3^c \cap \tilde{X}_4
\end{array}$$

- The information diagram can be displayed in two dimensions.

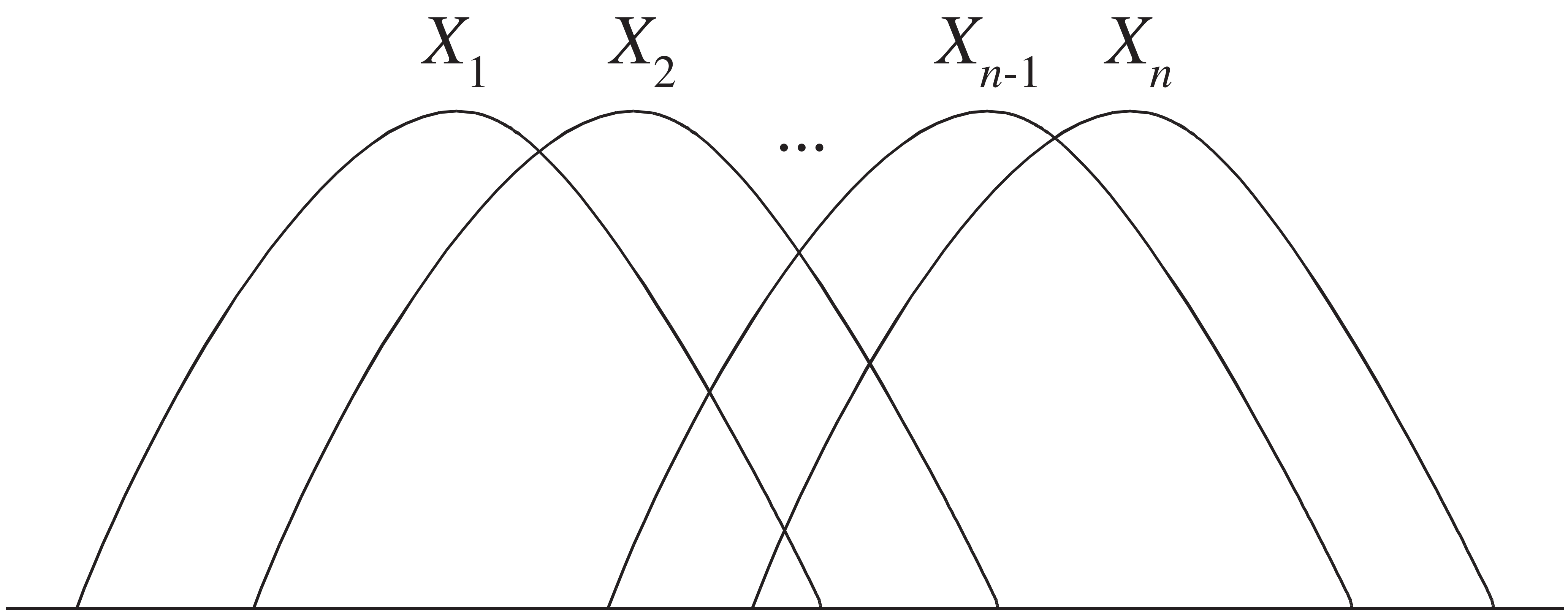
- The values of μ^* on the remaining atoms correspond to Shannon's information measures and hence are nonnegative. Thus, μ^* is a measure.


X_1
X_2
X_3
X_4

- For a general n, the information diagram can be displayed in two dimensions because certain atoms can be suppressed.

- The values of μ^* on the remaining atoms correspond to Shannon's information measures and hence are nonnegative. Thus, μ^* is a measure.

- See Ch. 12 for a detailed discussion in the context of Markov random field.



3.6 Examples of Applications

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, if $A \subset B$, then

    $$\mu^*(A) \le \mu^*(B)$$

    because

    $$\mu^*(A) \le \mu^*(A) + \mu^*(B - A) = \mu^*(B)$$

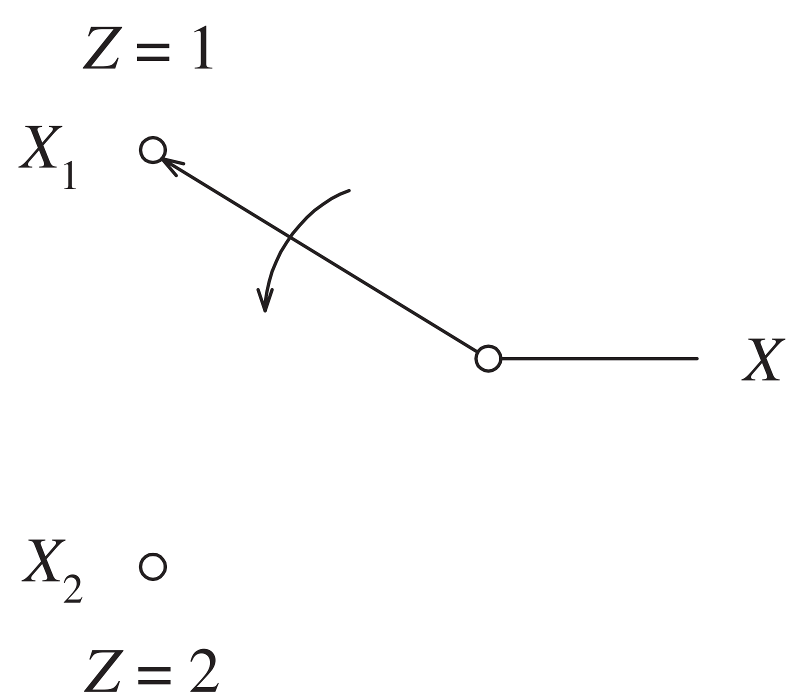
  - If μ^* is a signed measure, need to invoke the basic inequalities.

Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

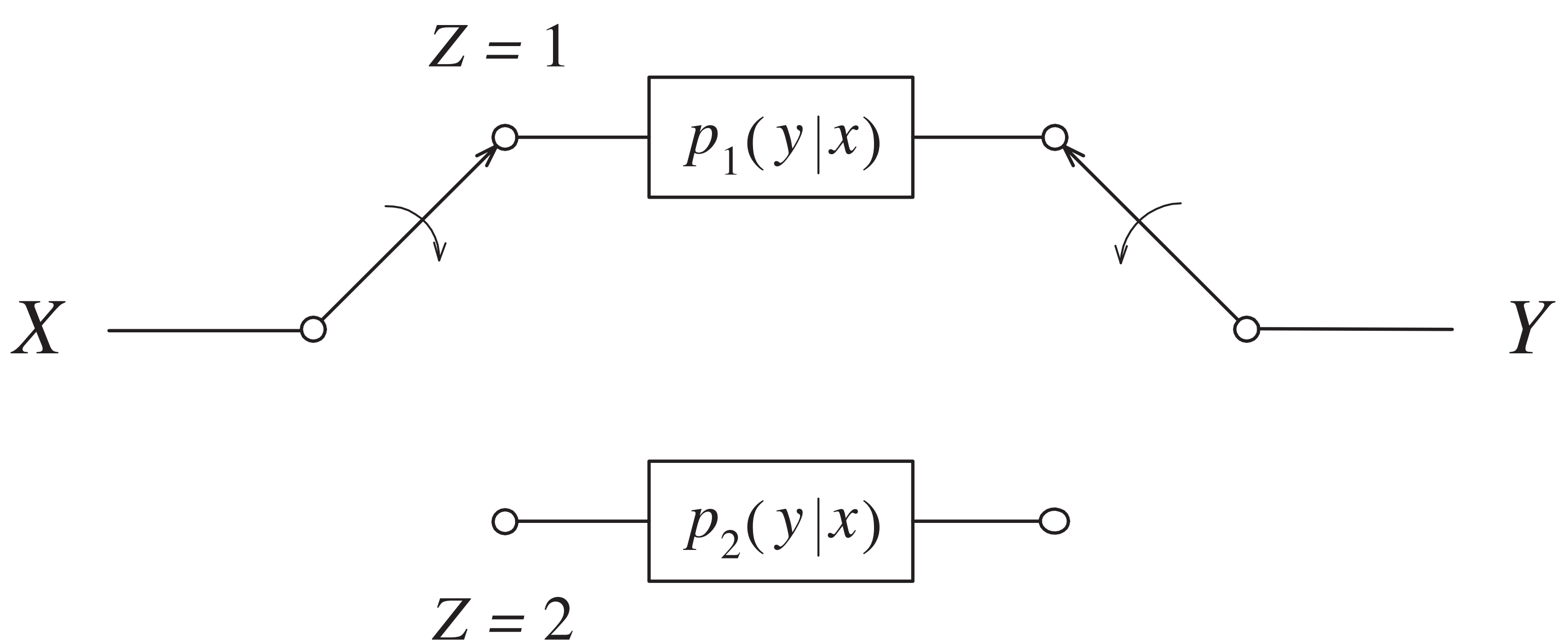
$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2).$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

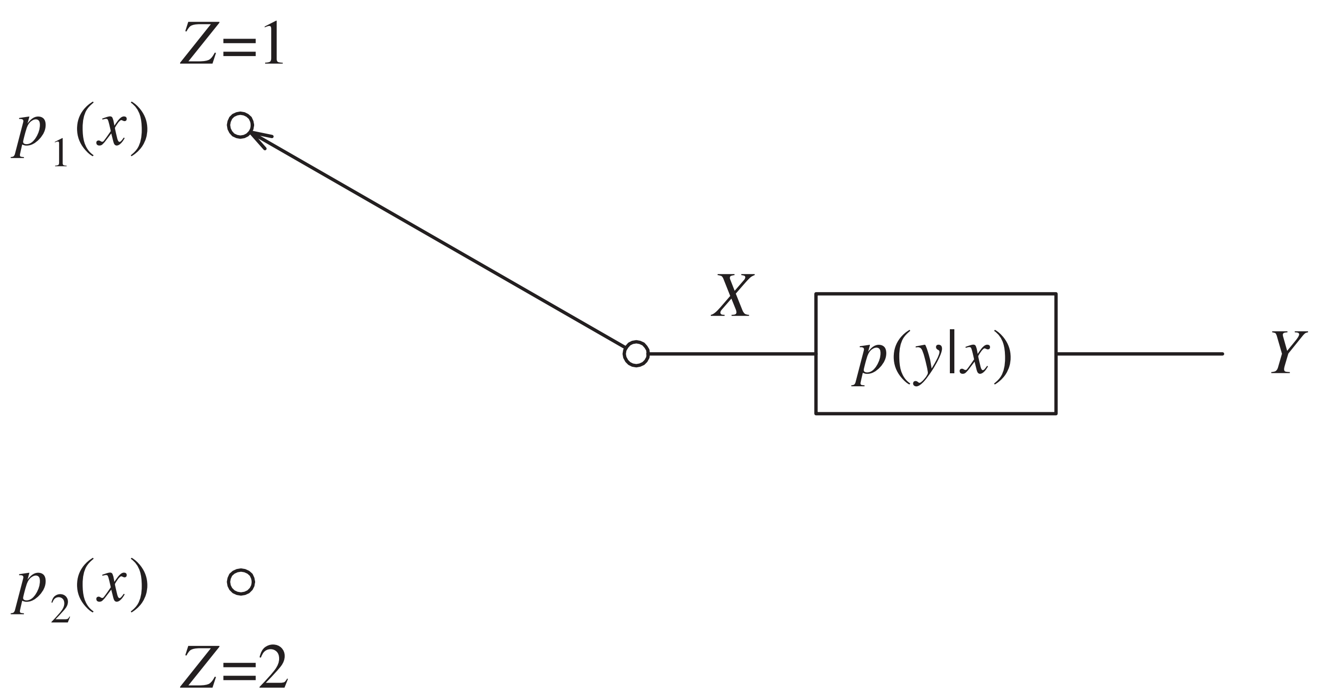


Setup: $I(X;Z) = 0$.

Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.



Setup: $Z \to X \to Y$.

Shannon's Perfect Secrecy Theorem

- X – plaintext
  Y – ciphertext
  Z – key
- Perfect Secrecy: $I(X;Y) = 0$
- Decipherability: $H(X|Y,Z) = 0$
- These requirement implies $H(Z) \geq H(X)$, i.e., the length of the key is at least the same as the length of the plaintext. Lower bound achievable by "one-time pad".
- Shannon (1949) gave a combinatorial proof.
- Can readily be proved by an information diagram.

Example 3.15 (Imperfect Secrecy Theorem) Let X be the plain text, Y be the cipher text, and Z be the key in a secret key cryptosystem. Since X can be recovered from Y and Z, we have

$$H(X|Y, Z) = 0.$$

Show that this constraint implies

$$I(X; Y) \geq H(X) - H(Z).$$

Remark Do not need to make these assumptions about the scheme:

- $H(Y|X, Z) = 0$
- $I(X; Z) = 0$

Example 3.17 (Data Processing Theorem) If $X \to Y \to Z \to T$, then

- $I(X;T) \leq I(Y;Z)$
- in fact

$$I(Y;Z) = I(X;T) + I(X;Z|T) + I(Y;T|X) + I(Y;Z|X,T)$$

Example 3.18 If $X \to Y \to Z \to T \to U$, then

$$H(Y) + H(T) = I(Z; X, Y, T, U) + I(X, Y; T, U) + H(Y|Z) + H(T|Z)$$

- Very difficult to discover without an information diagram.
- Instrumental in proving an outer bound for the multiple description problem.

Highlight of Ch. 12

- The I-Measure completely characterizes a class of Markov structures called full conditional independence.
- Markov random field is a special case.
- Markov chain is a special case of Markov random field.
- Analysis of these Markov structures becomes completely set-theoretic.

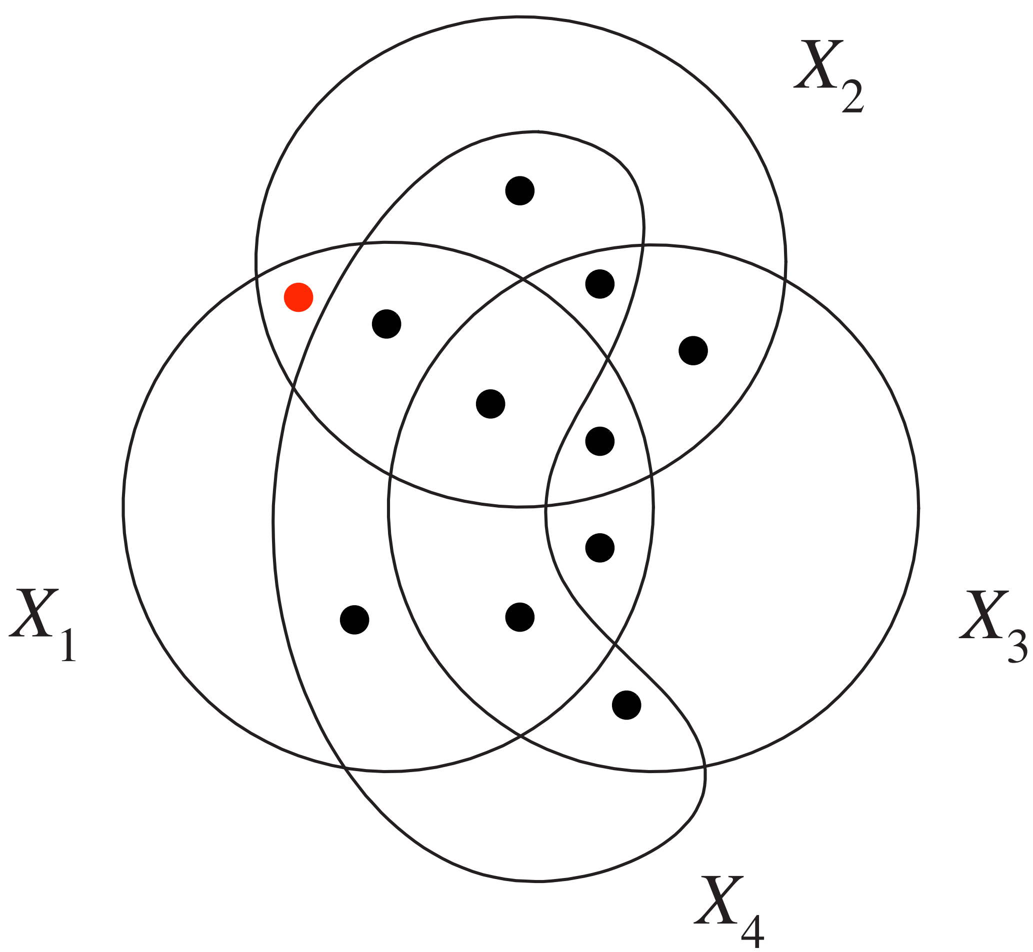
Example 12.22

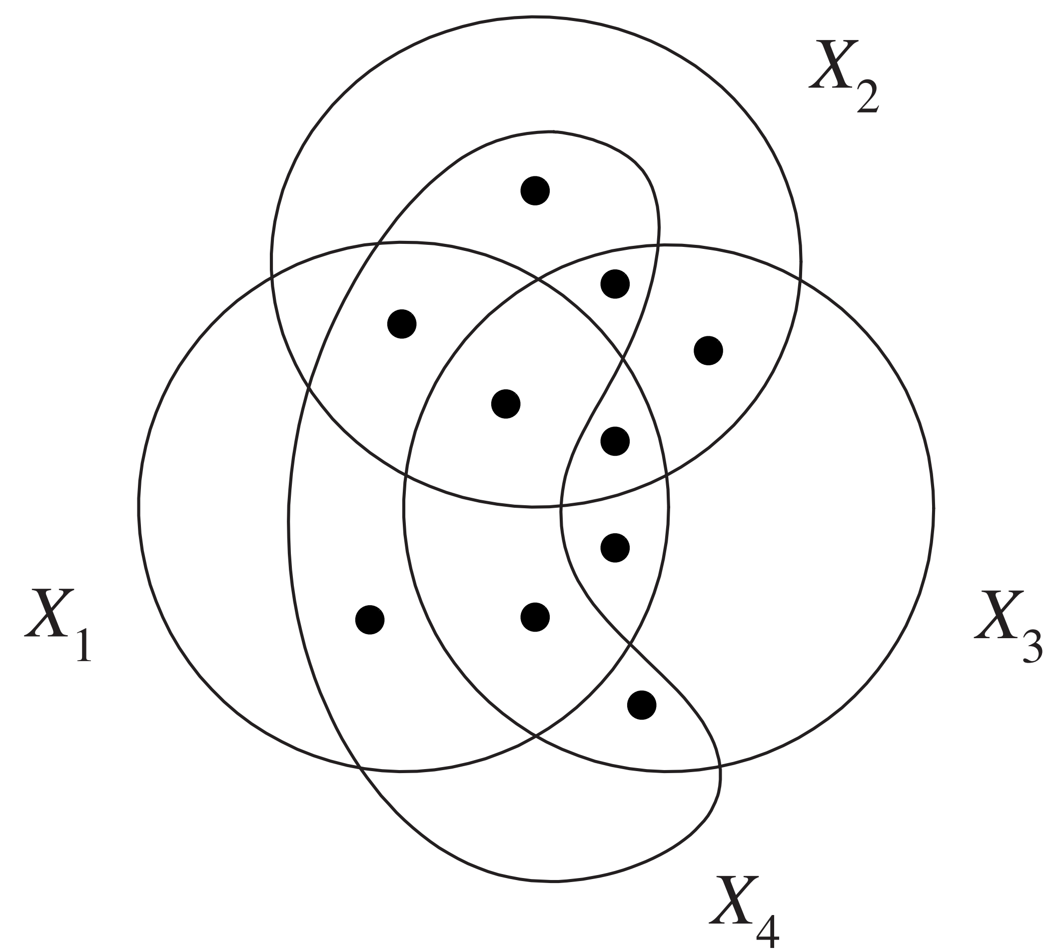
$$\Pi_1 : \left\{ \begin{array}{l} (X_1, X_2) \perp (X_3, X_4) \\ (X_1, X_3) \perp (X_2, X_4) \end{array} \right. \Rightarrow \Pi_2 : \left\{ \begin{array}{l} (X_1, X_2, X_3) \perp X_4 \\ (X_1, X_2, X_4) \perp X_3 \end{array} \right.$$

- Each (conditional) independency forces μ^* to vanish on the atoms in the corresponding set.
- E.g., $(X_1, X_2) \perp (X_3, X_4) \Leftrightarrow \mu^*$ vanishes on the atoms in $(\tilde{X}_1 \cup \tilde{X}_2) \cap (\tilde{X}_3 \cup \tilde{X}_4)$.

Analysis of Example 12.12

μ^* vanishes on atoms with a dot.





$\Pi_1 \quad \Rightarrow \quad \Pi_2$

Proving Information Inequalities

- Information inequalities that are implied by the basic inequalities are called Shannon-type inequalities.
- They can be proved by means of a linear program called ITIP (Information Theoretic Inequality Prover), developed on Matlab at CUHK (1996):

  http://user-www.ie.cuhk.edu.hk/~ITIP/

- A version running on C called Xitip was developed at EPFL (2007):

  http://xitip.epfl.ch/

- See Ch. 13 and 14 for discussion.

ITIP Examples

1. ```
   >> ITIP('H(XYZ) <= H(X) + H(Y) + H(Z)')
   True
   ```
2. ```
   >> ITIP('I(X;Z) = 0','I(X;Z|Y) = 0','I(X;Y) = 0')
   True
   ```
3. ```
   >> ITIP('X/Y/Z/T', 'X/Y/Z', 'Y/Z/T')
   Not provable by ITIP
   ```
4. ```
   >> ITIP('I(Z;U) - I(Z;U|X) - I(Z;U|Y) <=
    0.5 I(X;Y) + 0.25 I(X;ZU) + 0.25 I(Y;ZU)')
   Not provable by ITIP
   ```

- #4 is a so-called non-Shannon-type inequalities which is valid but not implied by the basic inequalities. See Ch. 15 for discussion.