Entropy and Mutual Information of the Conditional Independence Model

Nabeel Jawad Hassan
Department of Mathematics, College of Education for Pure Sciences
Nabeel79.Jawad@yahoo.com
Ahmed Abbas Mezaal
Department of Mathematics, College of Computer Science and Mathematics
aam_mb@yahoo.com

Abstract:
In this paper, we introduce the conditional independence model that included three elements every element is called conditional independence statement, or some times called network, those networks are defined by discrete random variables, we prove the conditional entropy function of the every element from this model. We also prove the conditional mutual information for every network, through out mesure the distance between the joint conditional probability density function and conditional probability density functions for every element in this model.

المستخلص:في هذا البحث, عرفنا نموذج الاستقلال الشرطي الذي يحتوي على ثلاث عناصر, نسمي كل عنصر من عناصر هذا النموذج بعبارة الاستقلال الشرطي وفي بعض الاحيان نسمي كل عنصر من النموذج بشبكة. هذه الشبكات عُرّفت بواسطة متغيرات عشوائية متقطعة. نحن برهنا دالة الشك لكل عنصر من عناصر هذا النموذج. كذلك برهنا المعلومات التبادلية الى كل شبكة من خلال قياس المسافة بين دالة كثافة الاحتمال الشرطية المشتركة ودوال كثافة الاحتمال الشرطية لكل عنصر في النموذج.

Keywords:Entropy function, Mutual information, conditional independence.

Introduction :

Entropy in its basic form is a measure of uncertainty rather than a measure of information. Specifically, the entropy of a random variable a measure of the uncertainty associated with that r.v. when the entropy of a random v. is large this means that the uncertainty as to the value of random is large and vice versa. The relative entropy gives measure of some think like the distance between two different p.d.f. . The conditional independence model is finite set from the conditional independence statements, Every conditional independence statement, can be represented as directed graph or indirected graph. A. J. Khinchine is spoke the basic ideas about the entropy in probability theory [1]. R. Gray is discuss the entropy function, and mutual information in [6]. T.Cover and J. Thomas, are studied the element concepts about the information theory in [9]. D. Geiger and C. Meek, both studied graphical models that included finite of the conditional independence statements in [2]. F. Matus, is introduced the conditional independence statement defined by four random variable in [3]. In this paper, we introduce conditional independence model $\{x_1 \coprod x_2 \coprod x_3/x_4,\ x_1 \coprod \{x_2, x_3\}/x_4,\ x_1 \coprod x_2/\{x_3, x_4\}\}$ every element from this model is network, the vertices are the random variables. In second section, we prove the entropy of every element in the conditional independence model. In third section, we prove the mutual information for every network , through out mesure destance between joint conditional probability density function and conditional probability density functions for every network in this model.

Definition (1.1) [5]:
Let $x = (x_1, x_2, \dots, x_n)$ be an m-dimensional random vector that takes its values in the $[m] = \prod_{i=1}^{n} m_i$, then the joint probability distribution has density function $f(x) = f(x_1, x_2, \dots, x_n)$.
Definition (1.2) [5]:
Let $A, B, C \subseteq [m]$ be pairwise disjoint. The random vector x_A is conditional independence of x_B given x_C if and only if : $f(x_A, x_B/x_C) = f(x_A/x_C) . f(x_B/x_C)$.
Remark (1.3) [5]:
The notation $x_A \coprod x_B/x_C$ is called conditional independence statement, and used to denoted the relationship that x_A is conditionally independentof x_B given x_C .
Definition (1.4):
Let $x = (x_1, x_2, x_3, x_4) = \{x_i : i \in \{1,2,3,4\}\}$ be a random vector with joint probability density function $f(x_1, x_2, x_3, x_4)$ with state space $[m] = \prod_{i=1}^{4} m_i$, where m_i is state space of x_i , then:
i) Let $\{1\},\{2\},\{3\},\{4\} \subseteq [m]$, be a pairwise disjoints, then x_1 is conditional independent x_2 and conditional independent x_3 given x_4 iff

$$p(x_1, {}^{x_2, x_3}/_{x_4}) = p({}^{x_1}/_{x_4}) . p({}^{x_2}/_{x_4}) . p({}^{x_3}/_{x_4})$$

ii) Let $\{1\},\{2,3\},\{4\} \subseteq [m]$, be a pairwise disjoints, then x_1 is conditional independent $\{x_2, x_3\}$ given x_4 iff: $p(x_1, x_2, x_3/x_4) = p(x_1/x_4) . p(x_2, x_3/x_4)$.
iii) Let $\{1\},\{2\},\{3,4\} \subseteq [m]$, be a pairwise disjoints, then x_1 is conditional independent x_2 given $\{x_3, x_4\}$ iff: $p(x_1, x_2/x_3, x_4) = p(x_1/x_3, x_4) . p(x_2/x_3, x_4)$.
Definition (1.5):
Let $x_1 \coprod x_2 \coprod x_3/x_4$, $x_1 \coprod \{x_2, x_3\} /x_4$, and $x_1 \coprod x_2/\{x_3, x_4\}$ are conditional independent statements, then the conditional independence model is

$$M = \{x_1 \coprod x_2 \coprod x_3 /x_4 , x_1 \coprod \{x_2, x_3\}/x_4 , x_1 \coprod x_2 / \{x_3, x_4\}\}.$$

Definition (1.6) [9]:
The entropy of discrete random variable x is defined by :

$$H(x) = -E(\log p(x)) = -\sum_{x\in A} p(x) . \log p(x)$$

where A is variable space of x .
Definition (1.7) [9]:
The joint entropy $H(x, y)$ of pair of discrete random variables, with a joint distribution $p(x, y)$is defined as:

$$H(x, y) = -E(\log p({}^{x}/_{y})) = -\sum_{x\in A}\sum_{y\in\beta} p(x, y) . \log p(x, y)$$

where A is variable space of x and β is variable space of y .
Definition (1.8) [9]:
The conditional entropy $H(x/y)$ is defined as:

$$H(x/y) = -E(\log p(x/y)) = -\sum_{x\in A}\sum_{y\in\beta} p(x, y) . \log p(x/y).$$

Definition (1.9) [9]:
The relative entropy between two probability density function $p(x)$, $q(x)$ is defined as:

$$D(p(x)\backslash\backslash q(x)) = \sum_{x\in A} p(x) . \log\frac{p(x)}{q(x)} = E(\log\frac{p(x)}{q(x)})$$

where A is variable space.

Definition (1.10) [9]:
Let x , y be r.v.s with joint probability density function, and marginal distributions p(x) and p(y). The mutual information I(x, y) is the relative entropy between the joint distribution and product distribution:

$$I(x,y) = D\big(p(x,y)\backslash\backslash p(x).\,p(y)\big) = \sum_{x\in A}\sum_{y\in\beta} p(x,y)\,.\log\frac{p(x,y)}{p(x).\,p(y)}.$$

2. Conditional Entropy

In this section, we prove entropy function for every network in the conditional independence model. We called conditional entropy, because, we prove entropy function for conditional independence statement.

Theorem (2.1):
Let $x = (x_1,\ x_2,\ x_3\ ,x_4)$ be a random vector, with joint probability density function $p(x_1,\ x_2,\ x_3, x_4)$ and marginal probability density function of x_4 ;$p(x_4)$, the entropy of conditional independence statement (CI.) $x_1 \coprod x_2 \coprod x_3/x_4$ is

$$H(x_1,x_2,x_3/x_4) = H(x_1/x_4) + H(x_2/x_4) + H(x_3/x_4)\ .$$

Proof:

$$H\big(x_1,{}^{x_2\ ,x_3}/_{x_4}\big) = -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1\ ,x_2,x_3,x_4)\log p(x_1,x_2,{}^{x_3}/_{x_4})$$

Since: $p(x_1,x_2,x_3/x_4) = p(x_1/x_4)\ .\,p(x_2/x_4)\cdot p(x_3/x_4)$

$$= -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1\ ,x_2,x_3,x_4)\cdot \log p({}^{x_1}/_{x_4})\cdot p({}^{x_2}/_{x_4})\cdot p({}^{x_3}/_{x_4})$$

$$= -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1\ ,x_2,x_3,x_4)\cdot \log p\big({}^{x_1}/_{x_4}\big) - \sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1,x_2\ ,x_3,x_4)$$

$$\cdot \log p\big({}^{x_2}/_{x_4}\big) - \sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1,x_2\ ,x_3,x_4)\cdot \log p({}^{x_3}/_{x_4})$$

Since: $p(x_1,\ x_2,\ x_3,x_4) = p(x_1\ ,x_2,x_3/x_4)\ .\,p(x_4) = p(x_1/x_4)\ .\,p(x_2/x_4).\,p(x_3\ /\ x_4)\cdot p(x_4)$

Then

$$= -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p({}^{x_1}/_{x_4})\cdot p({}^{x_2}/_{x_4})\cdot p({}^{x_3}/_{x_4})\cdot p(x_4).\log p\big({}^{x_1}/_{x_3}\big) -$$

$$\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p({}^{x_1}/_{x_4}).\,p\big({}^{x_2}/_{x_4}\big)\cdot p\big({}^{x_3}/_{x_4}\big)\cdot p(x_4)\,.\log\ p\big({}^{x_2}/_{x_3}\big) -$$

$$\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1\ /\ x_4)\cdot p(x_2\ /\ x_4)\cdot p(x_3\ /\ x_4)\cdot p(x_4)\cdot \log p(x_3\ /\ x_4)$$

$$= -\sum_{x_1} p\big({}^{x_1}/_{x_4}\big).\log p\big({}^{x_1}/_{x_4}\big)\ .\sum_{x_2} p\big({}^{x_2}/_{x_4}\big)\ .\sum_{x_3} p(x_3\ /\ x_4)\ \cdot\sum_{x_4} p(x_4) -$$

$$\sum_{x_2} p\left(x_2/x_4\right). \log p\left(x_2/x_4\right) . \sum_{x_1} p\left(x_1/x_4\right) . \sum_{x_3} p(x_3 / x_4) \cdot \sum_{x_4} p(x_4)$$

$$- \sum_{x_3} p(x_3 / x_4) \cdot \log p\left(x_3/x_4\right) \cdot \sum_{x_1} p(x_1 / x_4) \cdot \sum_{x_2} p(x_2 / x_4)$$

Since:

$$\sum_{x_2} p\left(x_2/x_4\right) = \sum_{x_3} p(x_3 / x_4) = \sum_{x_4} p(x_4) = \sum_{x_1} p(x_1 / x_4) = 1$$

We get:

$$= -\sum_{x_1} p\left(x_1/x_4\right). \log p\left(x_1/x_4\right) - \sum_{x_2} p\left(x_2/x_4\right) . \log\ p\left(x_2/x_4\right) - \sum_{x_3} p(x_3 / x_4) \cdot \log p(x_3 / x_4)$$

$$= H\left(x_1/x_4\right) + H\left(x_2/x_4\right) + H\left(x_3/x_4\right).$$

Theorem (2.2):

Let $x = (x_1, x_2, x_3, x_4)$ be a random vector, with joint probability density function$p(x_1, x_2, x_{3,}, x_4)$, and marginal probability density function of x_4 ; $p(x_4)$, the entropy of CI. $x_1 \coprod \{x_2, x_3\} / x_4$ is : $H(x_1, x_2, x_3/x_4) = H(x_1/x_4) + H(x_2, x_3/x_4)$.

Proof:

$$H\left(x_1, x_2, x_3/x_4\right) = -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1, x_2, x_3, x_4) . \log p(x_1, x_2, x_3/x_4)$$

Since: $p(x_1, x_2, x_3/x_4) = p(x_1/x_4) . p(x_2, x_3/x_4)$

Then

$$= -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1, x_2, x_3, x_4) . \log p\left(x_1/x_4\right). p\left(x_2, x_3/x_4\right)$$

Since: $p(x_1, x_2, x_3, x_4) = p(x_1, x_2, x_3/x_4) . p(x_4) = p(x_1/x_4) . p(x_2, x_3/x_4). p(x_4)$

We get:

$$= -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p\left(x_1/x_4\right) . p\left(x_2, x_3/x_4\right). p(x_4) . \log p\left(x_1/x_4\right) . p\left(x_2, x_3/x_4\right)$$

$$= -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p\left(x_1/x_4\right) . p\left(x_2, x_3/x_4\right). p(x_4) . \log p\left(x_1/x_4\right) -$$

$$\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p\left(x_1/x_4\right) . p\left(x_2, x_3/x_4\right). p(x_4) . \log p\left(x_2, x_3/x_4\right)$$

$$= -\sum_{x_1} p\left(x_1/x_4\right). \log p\left(x_1/x_4\right) . \sum_{x_2}\sum_{x_3} p\left(x_2, x_3/x_4\right) . \sum_{x_4} p(x_4) -$$

$$\sum_{x_2}\sum_{x_3} p\left({}^{x_2,x_3}/_{x_4}\right).\log p\left({}^{x_2,x_3}/_{x_4}\right).\sum_{x_1} p\left({}^{x_1}/_{x_4}\right).\sum_{x_4} p(x_4)$$

Since:

$$\sum_{x_2}\sum_{x_3} p\left({}^{x_2,x_3}/_{x_4}\right) = \sum_{x_4} p(x_4) = \sum_{x_1} p\left({}^{x_1}/_{x_4}\right) = 1$$

We get:

$$= -\sum_{x_1} p\left({}^{x_1}/_{x_4}\right).\log p\left({}^{x_1}/_{x_4}\right) - \sum_{x_2}\sum_{x_3} p\left({}^{x_2,x_3}/_{x_4}\right).\log p\left({}^{x_2,x_3}/_{x_4}\right)$$

$$= H\left({}^{x_1}/_{x_4}\right) + H\left({}^{x_2,\ x_3}/_{x_4}\right).$$

Theorem (2.3):

Let $x = (x_1, x_2, x_3, x_4)$ be a random vector, with joint probability density function $p(x_1, x_2, x_{3,}, x_4)$, and marginal probability density function of x_3, x_4 ; $p(x_3, x_4)$,

the entropy of CI. $x_1 \coprod x_2/\{x_3, x_4\}$ is :

$$H(x_1, x_2/x_3, x_4) = H(x_1/x_3, x_4) + H(x_2/x_3, x_4).$$

Proof:

$$H\left(x_1, {}^{x_2}/_{x_3}, x_4\right) = -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1, x_2, x_3, x_4).\log p(x_1, {}^{x_2}/_{x_3}, x_4)$$

Since: $p(x_1, x_2/x_3, x_4) = p(x_1/x_3, x_4) .\ p(x_2/x_3, x_4)$

Then

$$= -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1, x_2, x_3, x_4).\log p\left({}^{x_1}/_{x_3}, x_4\right).p\left({}^{x_2}/_{x_3}, x_4\right)$$

Since:

$p(x_1, x_2, x_3, x_4) = p(x_1, x_2/x_3, x_4).p(x_3, x_4) =$
$p(x_1/x_3, x_4).p(x_2/x_3, x_4).p(x_3, x_4)$

We get:

$$= -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p\left(x_1, {}^{x_2}/_{x_3}, x_4\right).p(x_3, x_4).\log p\left({}^{x_1}/_{x_3}, x_4\right).p\left({}^{x_2}/_{x_3}, x_4\right)$$

$$= -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p\left({}^{x_1}/_{x_3}, x_4\right).p\left({}^{x_2}/_{x_3}, x_4\right).p(x_3, x_4).\log p\left({}^{x_1}/_{x_3}, x_4\right).p\left({}^{x_2}/_{x_3}, x_4\right)$$

$$= -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p\left({}^{x_1}/_{x_3}, x_4\right).p\left({}^{x_2}/_{x_3}, x_4\right).p(x_3, x_4).\log p\left({}^{x_1}/_{x_3}, x_4\right) -$$

$$\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p\left({}^{x_1}/_{x_3}, x_4\right).p\left({}^{x_2}/_{x_3}, x_4\right).p(x_3, x_4).\log p\left({}^{x_2}/_{x_3}, x_4\right)$$

$$= -\sum_{x_1} p\left({}^{x_1}/_{x_3}, x_4\right).\log p\left({}^{x_1}/_{x_3}, x_4\right) .\sum_{x_2}\sum_{x_3} p(x_2, x_3) .\sum_{x_4} p\left({}^{x_2}/_{x_3}, x_4\right) -$$

$$\sum_{x_2} p\left({}^{x_2}/_{x_3}, x_4\right).\log p\left({}^{x_2}/_{x_3}, x_4\right).\sum_{x_1} p\left({}^{x_1}/_{x_3}, x_4\right) .\sum_{x_4}\sum_{x_3} p(x_3, x_4)$$

Since:

$$\sum_{x_2} p\left({}^{x_2}/_{x_3}, x_4\right) = \sum_{x_1} p\left({}^{x_1}/_{x_3}, x_4\right) = \sum_{x_3}\sum_{x_4} p(x_3, x_4) = 1$$

We get:

$$= -\sum_{x_1} p\left({}^{x_1}/_{x_3}, x_4\right).\log p\left({}^{x_2}/_{x_3}, x_4\right) - \sum_{x_2} p\left({}^{x_2}/_{x_3}, x_4\right)\log p\left({}^{x_2}/_{x_3}, x_4\right)$$

$$= H\left({}^{x_1}/_{x_3},\ x_4\right) + H\left({}^{x_2}/_{x_3}, x_4\right).$$

3 Conditional Mutual Information

In this section, we prove mutual information for every network in model through out mesure destance between joint conditional probability density function and conditional probability density functions for every network in the conditional independence model.We called conditional mutual information, because, we prove mutual information between the conditional probability density function.

Theorem (3.1):

Let$x = (x_1,\ x_2,\ x_3, x_4)$ be a random vector with joint probability density function$p(x_1,\ x_2,\ x_3, x_4)$, and marginal probability density function of x_4 ; $p(x_4)$, the mutual information between $p(x_1,\ x_2, x_3/x_4)$ and $p(x_1/x_4)\cdot p(x_2/x_4)\cdot p(x_3/x_4)$, is:

$$I(x_1,\ x_2, x_3/x_4) = H(x_1/x_4) + H(x_2/x_4) + H(x_3/x_4) - H(x_1,\ x_2, x_3/x_4).$$

Proof:

$$I\left(x_1,\ {}^{x_2,x_3}/_{x_4}\right) = D\left[\ p\left(x_1,\ {}^{x_2,x_3}/_{x_4}\right)\backslash\backslash\ p\left({}^{x_1}/_{x_4}\right).p\left({}^{x_2}/_{x_4}\right)\cdot p\left({}^{x_3}/_{x_4}\right)\right]$$

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1, x_2, x_3, x_4).\log\frac{p\left(x_1,\ {}^{x_2,x_3}/_{x_4}\right)}{p\left({}^{x_1}/_{x_4}\right)\ \cdot\ p\left({}^{x_2}/_{x_4}\right)\cdot p\left({}^{x_3}/_{x_4}\right)}$$

Since: $p(x_1,\ x_2,\ x_3, x_4) = p(x_1,\ x_2, x_3/x_4).\ p(x_4)$

Then

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1, x_2, x_3\ /\ x_4).\ p(x_4).\log\frac{p\left(x_1,\ {}^{x_2,x_3}/_{x_4}\right)}{p\left({}^{x_1}/_{x_4}\right)\cdot\ p\left({}^{x_2}/_{x_4}\right)\cdot p\left({}^{x_3}/_{x_4}\right)}$$

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3} p\left(x_1, x_2, {}^{x_3}/_{x_4}\right).\log\frac{p\left(x_1,\ {}^{x_2,x_3}/_{x_4}\right)}{p\left({}^{x_1}/_{x_4}\right).\ p\left({}^{x_2}/_{x_4}\right)\cdot p(x_3\ /\ x_4)}\cdot\sum_{x_4} p(x_4)$$

Since :

$$\sum_{x_4} p(x_4) = 1$$

Then

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, x_2, x_3 / x_4).\log p\left(x_1, {}^{x_2, x_3}/_{x_4}\right)$$

$$- \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, x_2, {}^{x_3}/_{x_4}).\log p\left({}^{x_1}/_{x_4}\right).p\left({}^{x_2}/_{x_4}\right)\cdot p\left({}^{x_3}/_{x_4}\right)$$

Since $p(x_1, x_2, x_3/x_4) = p(x_1/x_4)\cdot p(x_2/x_4)\cdot p(x_3 / x_4)$
Then:

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, x_2, {}^{x_3}/_{x_4}).\log p\left(x_1, {}^{x_2, x_3}/_{x_4}\right) - \sum_{x_1} p\left({}^{x_1}/_{x_4}\right)\cdot \log p\left({}^{x_1}/_{x_4}\right)$$

$$\cdot \sum_{x_2} p\left({}^{x_2}/_{x_4}\right)\cdot \sum_{x_3} p(x_3 / x_4) -$$

$$\sum_{x_2} p(x_2/x_4).\log p(x_2/x_4)\cdot \sum_{x_1} p(x_1 / x_4)\cdot \sum_{x_3} p(x_3/x_4) -$$

$$\sum_{x_3} p(x_3/x_4)\cdot \log p(x_3/x_4)\cdot \sum_{x_1} p(x_1 / x_4)\cdot \sum_{x_2} p(x_2/x_4)$$

Since:

$$\sum_{x_1} p(x_1 / x_4) = \sum_{x_2} p(x_2 / x_4) = \sum_{x_3} p(x_3 / x_4) = 1$$

Then:

$$= H\left({}^{x_1}/_{x_4}\right) + H\left({}^{x_2}/_{x_4}\right) + H\left({}^{x_3}/_{x_4}\right) - H\left(x_1, {}^{x_2, x_3}/_{x_4}\right).$$

Theorem (3.2):

Let$x = (x_1, x_2, x_3, x_4)$ be a random vector with joint probability density function$p(x_1, x_2, x_3, x_4)$, and marginal probability density function of x_4 ; $p(x_4)$, the mutual information between $p(x_1, x_2, x_3/x_4)$ and $p(x_1/x_4).p(x_2, x_3/x_4)$ is:

$$I(x_1, x_2, x_3/x_4) = H(x_1/x_4) + H(x_2, x_3/x_4) - H(x_1, x_2, x_3/x_4).$$

Proof:

$$I\left(x_1, {}^{x_2, x_3}/_{x_4}\right) = D\left[p\left(x_1, {}^{x_2, x_3}/_{x_4}\right)//p\left({}^{x_1}/_{x_4}\right).p\left({}^{x_2, x_3}/_{x_4}\right)\right]$$

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1, x_2, x_3, x_4).\log \frac{p\left(x_1, {}^{x_2, x_3}/_{x_4}\right)}{p\left({}^{x_1}/_{x_4}\right).\ p\left({}^{x_2, x_3}/_{x_4}\right)}$$

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_4).p\left(x_1, {}^{x_2, x_3}/_{x_4}\right).\log \frac{p\left(x_1, {}^{x_2, x_3}/_{x_4}\right)}{p\left({}^{x_1}/_{x_4}\right).\ p\left({}^{x_2, x_3}/_{x_4}\right)}$$

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3} p\left(x_1, {}^{x_2, x_3}/_{x_4}\right).\log \frac{p\left(x_1, {}^{x_2, x_3}/_{x_4}\right)}{p\left({}^{x_1}/_{x_4}\right).\ p\left({}^{x_2, x_3}/_{x_4}\right)}\ .\sum_{x_4} p(x_4)$$

Since:

$$\sum_{x_4} p(x_4) = 1$$

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, {}^{x_2,\, x_3}/_{x_4}) \,.\log \frac{p(x_1, {}^{x_2,\, x_3}/_{x_4})}{p({}^{x_1}/_{x_4})\,.\; p({}^{x_2, x_3}/_{x_4})}$$

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, {}^{x_2,\, x_3}/_{x_4}) \,.\log p(x_1, {}^{x_2,\, x_3}/_{x_4}) -$$

$$\sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, {}^{x_2,\, x_3}/_{x_4}) \,.\log p({}^{x_1}/_{x_4}).\, p({}^{x_2,\, x_3}/_{x_4})$$

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, {}^{x_2,\, x_3}/_{x_4}) \,.\log p(x_1, {}^{x_2,\, x_3}/_{x_4}) -$$

$$\sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, {}^{x_2,\, x_3}/_{x_4}).\log p({}^{x_1}/_{x_4}) -$$

$$\sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, {}^{x_2,\, x_3}/_{x_4}).\log p({}^{x_2,\, x_3}/_{x_4})$$

Since: $p(x_1,\, x_2,\, x_3/x_4) = p(x_1/x_4).\, p(x_2,\, x_3/x_4)$

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, {}^{x_2,\, x_3}/_{x_4}) \,.\log p(x_1, {}^{x_2,\, x_3}/_{x_4}) -$$

$$\sum_{x_1} p({}^{x_1}/_{x_4}) \,.\log p({}^{x_1}/_{x_4}).\sum_{x_2}\sum_{x_3} p({}^{x_2,\, x_3}/_{x_4}) -$$

$$\sum_{x_2}\sum_{x_3} p({}^{x_2,\, x_3}/_{x_4})\,.\; \log p({}^{x_2,\, x_3}/_{x_4}).\sum_{x_1} p({}^{x_1}/_{x_4})$$

Since:

$$\sum_{x_2}\sum_{x_3} p({}^{x_2,\, x_3}/_{x_4}) = \sum_{x_1} p({}^{x_1}/_{x_4}) = 1$$

Then

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, {}^{x_2,\, x_3}/_{x_4}) \,.\log p(x_1, {}^{x_2,\, x_3}/_{x_4}) -$$

$$\sum_{x_1} p({}^{x_1}/_{x_4}).\log p({}^{x_1}/_{x_4}) - \sum_{x_2}\sum_{x_3} p({}^{x_2,\, x_3}/_{x_4})\,.\; \log p({}^{x_2,\, x_3}/_{x_4})$$

$$= -H(x_1, {}^{x_2,\, x_3}/_{x_4}) + H({}^{x_1}/_{x_4}) + H({}^{x_2,\, x_3}/_{x_4})$$

$$= H({}^{x_1}/_{x_4}) + H({}^{x_2,\, x_3}/_{x_4}) - H(x_1, {}^{x_2,\, x_3}/_{x_4}).$$

Theorem (3.3):

Let$x = (x_1,\, x_2,\, x_3,\, x_4)$ be a random vector with joint probability density function$p(x_1,\, x_2,\, x_3,\, x_4)$, and marginal probability density function of x_3, x_4 ; $p(x_3, x_4)$. The mutual information between $p(x_1,\, x_2/x_3,\, x_4)$ and $p(x_1/x_3\,,\, x_4).\, p(x_2/x_3,\, x_4)$ is:

$$I(x_1, x_2/x_3, x_4) = H(x_1/x_3, x_4) + H(x_2/x_3, x_4) - H(x_1, x_2/x_3, x_4).$$

Proof:

$$I(x_1, {}^{x_2}/_{x_3, x_4}) = D[\, p(x_1, {}^{x_2}/_{x_3, x_4}) // p({}^{x_1}/_{x_3, x_4}).\, p({}^{x_2}/_{x_3, x_4})]$$

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1, x_2, x_3, x_4).\log \frac{p(x_1, {}^{x_2}/_{x_3, x_4})}{p({}^{x_1}/_{x_3, x_4}).\ p({}^{x_2}/_{x_3, x_4})}$$

Since:

$$p(x_1, x_2, x_3, x_4) = p(x_1, {}^{x_2}/_{x_3, x_4}).\, p(x_3, x_4).$$

Then

$$= \sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_3, x_4).\, p(x_1, {}^{x_2}/_{x_3, x_4}).\log \frac{p(x_1, {}^{x_2}/_{x_3, x_4})}{p({}^{x_1}/_{x_3, x_4}).\ p({}^{x_2}/_{x_3, x_4})}$$

$$= \sum_{x_1}\sum_{x_2} p(x_1, {}^{x_2}/_{x_3, x_4}).\log \frac{p(x_1, {}^{x_2}/_{x_3, x_4})}{p({}^{x_1}/_{x_3, x_4}).\ p({}^{x_2}/_{x_3, x_4})}\ .\sum_{x_3}\sum_{x_4} p(x_3, x_4)$$

Since:

$$\sum_{x_3}\sum_{x_4} p(x_3, x_4) = 1$$

Then

$$= \sum_{x_1}\sum_{x_2} p(x_1, {}^{x_2}/_{x_3, x_4}).\log \frac{p(x_1, {}^{x_2}/_{x_3, x_4})}{p({}^{x_1}/_{x_3, x_4}).\ p({}^{x_2}/_{x_3, x_4})}$$

$$= \sum_{x_1}\sum_{x_2} p(x_1, {}^{x_2}/_{x_3, x_4}).\log p(x_1, {}^{x_2}/_{x_3, x_4}) -$$

$$\sum_{x_1}\sum_{x_2} p(x_1, {}^{x_2}/_{x_3, x_4}).\log p({}^{x_1}/_{x_3, x_4}).\ p({}^{x_2}/_{x_3, x_4})$$

$$= \sum_{x_1}\sum_{x_2} p(x_1, {}^{x_2}/_{x_3, x_4}).\log p(x_1, {}^{x_2}/_{x_3, x_4}) -$$

$$\sum_{x_1}\sum_{x_2} p(x_1, {}^{x_2}/_{x_3, x_4}).\log p({}^{x_1}/_{x_3, x_4}) - \sum_{x_1}\sum_{x_2} p(x_1, {}^{x_2}/_{x_3, x_4}).\log\ p({}^{x_2}/_{x_3, x_4})$$

Since:

$$p(x_1, {}^{x_2}/_{x_3, x_4}) = p({}^{x_1}/_{x_3, x_4}).\, p({}^{x_2}/_{x_3, x_4}).$$

Then

$$= \sum_{x_1}\sum_{x_2} p(x_1, {}^{x_2}/_{x_3, x_4}).\log p(x_1, {}^{x_2}/_{x_3, x_4}) -$$

$$\sum_{x_1}\sum_{x_2} p({}^{x_1}/_{x_3, x_4}).\, p({}^{x_2}/_{x_3, x_4}).\log p({}^{x_1}/_{x_3, x_4}) -$$

$$\sum_{x_1}\sum_{x_2} p\left({}^{x_1}/_{x_3,\ x_4}\right).\, p\left({}^{x_2}/_{x_3,\ x_4}\right).\log\ p\left({}^{x_2}/_{x_3,\ x_4}\right)$$

$$= \sum_{x_1}\sum_{x_2} p\left(x_1,\ {}^{x_2}/_{x_3,\ x_4}\right).\log p\left(x_1,\ {}^{x_2}/_{x_3,\ x_4}\right) -$$

$$\sum_{x_1} p\left({}^{x_1}/_{x_3,\ x_4}\right).\log p\left({}^{x_1}/_{x_3,\ x_4}\right)\sum_{x_2} p\left({}^{x_2}/_{x_3,\ x_4}\right) -$$

$$\sum_{x_2} p\left({}^{x_2}/_{x_3,\ x_4}\right).\log p\left({}^{x_2}/_{x_3,\ x_4}\right)\sum_{x_1} p\left({}^{x_1}/_{x_3,\ x_4}\right)$$

Since:

$$\sum_{x_1} p\left({}^{x_1}/_{x_3,\ x_4}\right) = \sum_{x_2} p\left({}^{x_2}/_{x_3,\ x_4}\right) = 1\ .$$

Then

$$= \sum_{x_1}\sum_{x_2} p(x_1,\ x_2/x_3,\ x_4).\log p(x_1,\ x_2/x_3,\ x_4)$$

$$- \sum_{x_1} p(\ x_1/x_3,\ x_4).\log p(\ x_1/x_3,\ x_4) -$$

$$\sum_{x_2} p\left({}^{x_2}/_{x_3,\ x_4}\right).\log p\left({}^{x_2}/_{x_3,\ x_4}\right)$$

$$= H\left({}^{x_1}/_{x_3}\,,\ x_4\right) + H\left({}^{x_2}/_{x_3,\ x_4}\right) - H\left(x_1,\ {}^{x_2}/_{x_3,\ x_4}\right).$$

References:

[1] A. J. Khinchine: The entropy concept in probability theorem, Uspekhi Mathematicheskikh Nauk, 1953.
[2] D. Geiger and C. Meek: Graphical models and exponential families, proceedings of the fourteenth.
[3] F. Matus: Conditional independence among four random variables, Final conclusion. Combin. Probab. Comput., 8 (1999), 269-276.
[4] L. D. Garcia, M. Stillman, and B. Sturmfels: Algebraic geometry of Bayesian networks. Journal of Symbolic Computation, (2004).
[5] M. Drton, B. Sturmfels, and S. Sullivant: Lectures on Algebraic Statistics, (2008).
[6] M. Studeny: On Mathematical Description of probabilistic Conditional Independence Structures, Dr. Sc. Thesis, Prague, May (2001).
[7] R. Gray: Entropy and Information Theory. Electrical Engineering Department, Stanford University, Springer-Verlag, New york, (2009).
[8] S. L. Lauritzen: Graphical Models, Oxford University Press, 1996.
[9] T. Cover, and J. Thomas: Elements of Information Theory, Wiley Interscience, 1991.