ΘΕΩΡΙΑ ΠΛΗΡΟΦΟΡΙΑΣ

Κεφάλαιο 2 : Πληροφορία και Εντροπία

Κώστας Μαλιάτσος

Πανεπιστήμιο Αιγαίου, Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

Περιεχόμενα

- Πιθανότητες
- Πληροφορία
- Μέτρο πληροφορίας
- Μέση πληροφορία ή Εντροπία
- Από κοινού εντροπία
- Υπό συνθήκη εντροπία
- Αμοιβαία πληροφορία

Πιθανότητες (1)

- **Εκβάσεις ή ενδεχόμενα ή δειγματικά σημεία:**

λέγονται τα ατομικά αδιαίρετα αποτελέσματα ενός πειράματος, όπως στην περίπτωση του ζαριού τα 1,2,3,4,5,6.

- **Δειγματικός χώρος (sample space):**

λέγεται το σύνολο των πιθανών αποτελεσμάτων ενός τυχαίου πειράματος $\Omega = \{\omega_1, \omega_2, \ldots \omega_n\}$ π.χ. στο ζάρι $\Omega = \{1,2,3,4,5,6\}$

- **Γεγονός:** λέγεται κάθε υποσύνολο Α του δειγματικού χώρου, δηλαδή μια συλλογή εκβάσεων ή απλών ενδεχομένων ή δειγματικών σημείων.
- ορίζεται ως **βέβαιο γεγονός** το σύνολο του δειγματοχώρου, αφού θα συμβαίνει πάντα, και ως **μηδενικό γεγονός** το σύνολο που δεν περιέχει κανένα αποτέλεσμα και συνεπώς δεν θα συμβεί ποτέ.
- **Αν θεωρηθεί ότι ένα γεγονός Α αποτελείται από ν δειγματικά σημεία και ότι όλα τα σημεία του δειγματικού χώρου είναι Ν και ισοπίθανα, τότε ορίζεται ως πιθανότητα του Α ο λόγος ν / Ν .**

Πιθανότητες (2)

- Η **πιθανότητα** $P(A)$ ικανοποιεί τα ακόλουθα αξιώματα της θεωρίας πιθανοτήτων:

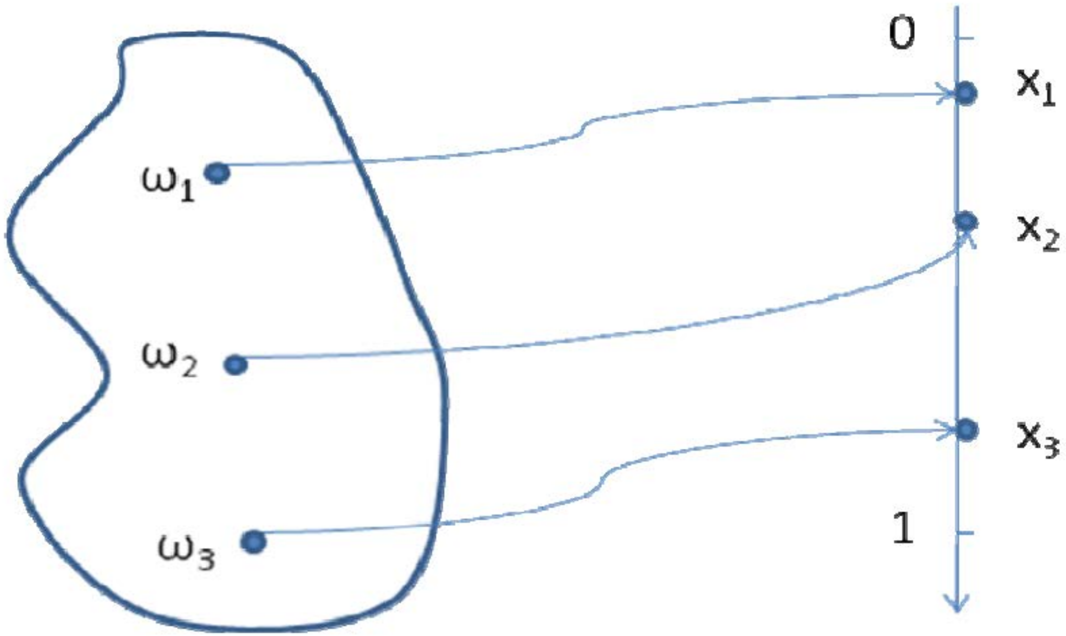
$$0 \le P(A) \le 1$$

  - $P[\varnothing] = 0$
  - $P[A^C] = 1 - P[A]$
  - $P[A \cup B] = P[A] + P[B] - P[A \cap B]$
  - Αν $A \subset B$ τότε $P[A] \le P[B]$

- Για κάθε δύο αποκλειστικά αμοιβαία (mutually exclusive) γεγονότα A και B ($A \cap B = \varnothing$) ισχύει: $P(A \cup B) = P(A) + P(B)$

Πιθανότητες (3)

- **Τυχαία μεταβλητή** είναι μια μονοσήμαντη συνάρτηση με πεδίο ορισμού ένα δειγματικό χώρο Ω και πεδίο τιμών ένα υποσύνολο των πραγματικών αριθμών.
- Ο ορισμός της τυχαίας μεταβλητής ως αντιστοίχιση στοιχειωδών γεγονότων ενός δειγματοχώρου με σημεία στον άξονα των πραγματικών αριθμών



Πιθανότητες (4)

- **Τυχαία μεταβλητή** είναι μια μονοσήμαντη συνάρτηση με πεδίο ορισμού ένα δειγματικό χώρο Ω και πεδίο τιμών ένα υποσύνολο των πραγματικών αριθμών.
- Ο ορισμός της τυχαίας μεταβλητής ως αντιστοίχιση στοιχειωδών γεγονότων ενός δειγματοχώρου με σημεία στον άξονα των πραγματικών αριθμών
- Μία τυχαία μεταβλητή λέγεται διακριτή αν το σύνολο των τιμών της είναι πεπερασμένο ή απείρως αριθμήσιμο
- Οι συνεχείς τυχαίες μεταβλητές αντιστοιχούν σε συνεχείς δειγματικούς χώρους.

Πιθανότητες (5)

- **Συνάρτηση μάζας πιθανότητας (Probability Mass Function)**

$$P(X = x_i) = p_i$$

και το σύνολο των πιθανοτήτων αυτών είναι $P = \{p_1, p_2, \dots p_n\}$

- **Ιδιότητες:**

$p(x_i) \geq 0$, για κάθε i

$$\sum_{i}^{n} p(x_i) = 1$$

- Η **συνάρτηση κατανομής αθροιστικής πιθανότητας (cumulative distribution function)** μιας διακριτής τυχαίας μεταβλητής *X* δίνεται από την παρακάτω σχέση:

$F(X \leq x) = \sum_{x_i \leq x} p(x_i)$, για κάθε $x \in (-\infty, \infty)$

Πιθανότητες (6)

- Αντίστοιχα, η συνάρτηση κατανομής μιας συνεχούς τυχαίας μεταβλητής δίνεται από:

$$F(X \le x) = P[X \in (-\infty, x)] = \int_{-\infty}^{x} f(y)dy \text{, για κάθε } x \in (-\infty, \infty)$$

- Η μη αρνητική συνάρτηση *f* (*x*) καλείται **συνάρτηση πυκνότητας πιθανότητας της συνεχούς τυχαίας μεταβλητής** *X* . Για τη συνάρτηση πυκνότητας πιθανότητας ισχύουν οι παρακάτω ιδιότητες:

$$\int_{B} f(x)dx = P(X \in B)$$

$$0 \le F(X \le x) \le 1 \text{, για κάθε } x$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

- Η συνάρτηση κατανομής είναι μη φθίνουσα, δηλαδή αν a≤b,

  F(*X* ≤ a) ≤ *F*(*X* ≤ b)

Πιθανότητες (7)

- Ας εξετάσουμε τώρα το πείραμα (X,Y) με δειγματικό χώρο το σύνολο των συνδυασμών (x, y) .
- Ορίζουμε ως **συνάρτηση συνδυασμένης πιθανότητας μάζας** την $p(x_i, y_j) = p_{ij} = P(X = x_i, Y = y_j)$

  που δίνει την πιθανότητα $X = x_i$ και $Y = y_j$
- Από τη συνάρτηση συνδυασμένης πιθανότητας μάζας μπορούν να υπολογιστούν οι **συναρτήσεις ακραίας πιθανότητας μάζας**

$$p(x_i) = \sum_{j=1}^{m} p_{ij} \quad \text{και} \quad p(y_j) = \sum_{i=1}^{n} p_{ij}$$

Πιθανότητες (8)

- Η **υπό συνθήκη πιθανότητα** είναι ένας άλλος τύπος πιθανότητας, που προκύπτει, όταν το αποτέλεσμα ενός πειράματος *Y* αποτελεί τη συνθήκη για ένα άλλο πείραμα *X*.
- Η συνάρτηση υπό συνθήκη πιθανότητας μάζας $p(x_i / y_j)$ που δίνει την πιθανότητα $X=x_i$ δεδομένου του $Y=y_i$, ορίζεται ως:

$$p(x_i / y_j) = \frac{p(x_i, y_j)}{p(y_j)}, \text{ εφόσον } p(y_j) > 0$$

Αντίστοιχα:

$$p(y_j / x_i) = \frac{p(x_i, y_j)}{p(x_i)}, \text{ εφόσον } p(x_i) > 0$$

Ακολούθως:

$$p(x_i, y_j) = p(x_i / y_j) p(y_j) = p(y_j / x_i) p(x_i)$$

Πιθανότητες (9)

- Θεώρημα του Bayes

$$p(x_i / y_j) = \frac{p(x_i, y_j)}{p(y_j)} = \frac{p(y_j / x_i)p(x_i)}{p(y_j)} = \frac{p(y_j / x_i)p(x_i)}{\sum_{i=1}^{n} p(y_j, x_i)} = \frac{p(y_j / x_i)p(x_i)}{\sum_{i=1}^{n} p(y_j / x_i)p(x_i)}$$

Πληροφορία

- Πληροφορία = νέα γνώση
  - Αποκτάται μέσω ακοής, όρασης, κλπ.
- Η πηγή πληροφορίας παράγει εξόδους που δεν είναι γνωστές στο δέκτη εκ των προτέρων
  - Εάν είχαμε τη δυνατότητα να τις προβλέψουμε δεν θα υπήρχε λόγος μετάδοσης.
- Τι είναι πληροφορία;
  - Ποιοτική περιγραφή
  - Απαιτείται όμως και ποσοτική περιγραφή

Πληροφορία

- Πηγές πληροφορίας με διακριτό αλφάβητο
- Αλφάβητο Διακριτής Πηγής:

$$\boxed{\Phi = \left\{ s_0, s_1, s_2 \right\}}$$

- Παράδειγμα: ο καιρός στην Ελλάδα κάθε 15 Αυγούστου
  - s_0: χιόνι
  - s_1: βροχή
  - s_2: λιακάδα
- Πότε δίνεται περισσότερη πληροφορία;
  - όταν τυχαίνει το σύμβολο s_0 ή το s_2;
  - με τι σχετίζεται η πληροφορία που φέρει κάθε σύμβολο;

Μέτρο Πληροφορίας

- Η πληροφορία ενός συμβόλου είναι μια φθίνουσα συνάρτηση της πιθανότητας εμφάνισης
- Μικρή αλλαγή στην πιθανότητα → μικρή αλλαγή στην πληροφορία (συνεχής συνάρτηση)
- Έστω ότι συνδυάζω δύο πηγές και φτιάχνω μια τρίτη:
  - $Φ_1$: καιρός στην Ελλάδα
  - $Φ_2$: δείκτης NASDAQ
  - Είναι ανεξάρτητες, τότε η πληροφορία της σύνθετης πηγής είναι το άθροισμα των πληροφοριών των δύο πηγών
  - πιθανότητα σύνθετου συμβόλου

Ιδιότητες Πληροφορίας

- Όταν η πιθανότητα να συμβεί ένα γεγονός είναι μονάδα, τότε η ποσότητα της μεταφερόμενης πληροφορίας είναι μηδενική

Όταν $P_A = 1$ τότε $I_A = 0$

- Η πληροφορία ενός γεγονότος είναι μη αρνητικό μέγεθος αφού ισχύει $I_A \geq 0$ και $0 \leq P_A \leq 1$

- Όσο πιο απίθανο είναι να συμβεί ένα γεγονός, τόσο περισσότερη πληροφορία λαμβάνουμε από την πραγματοποίηση του, δηλαδή

$$P_A \leq P_B \quad \text{τότε} \quad I_A \geq I_B$$

- Η συνολική ποσότητα πληροφορίας δύο ανεξάρτητων γεγονότων είναι το άθροισμα των επιμέρους μέτρων πληροφορίας των γεγονότων

$$I_{AB} = I_A + I_B \qquad P_{AB} = P_A \cdot P_B$$

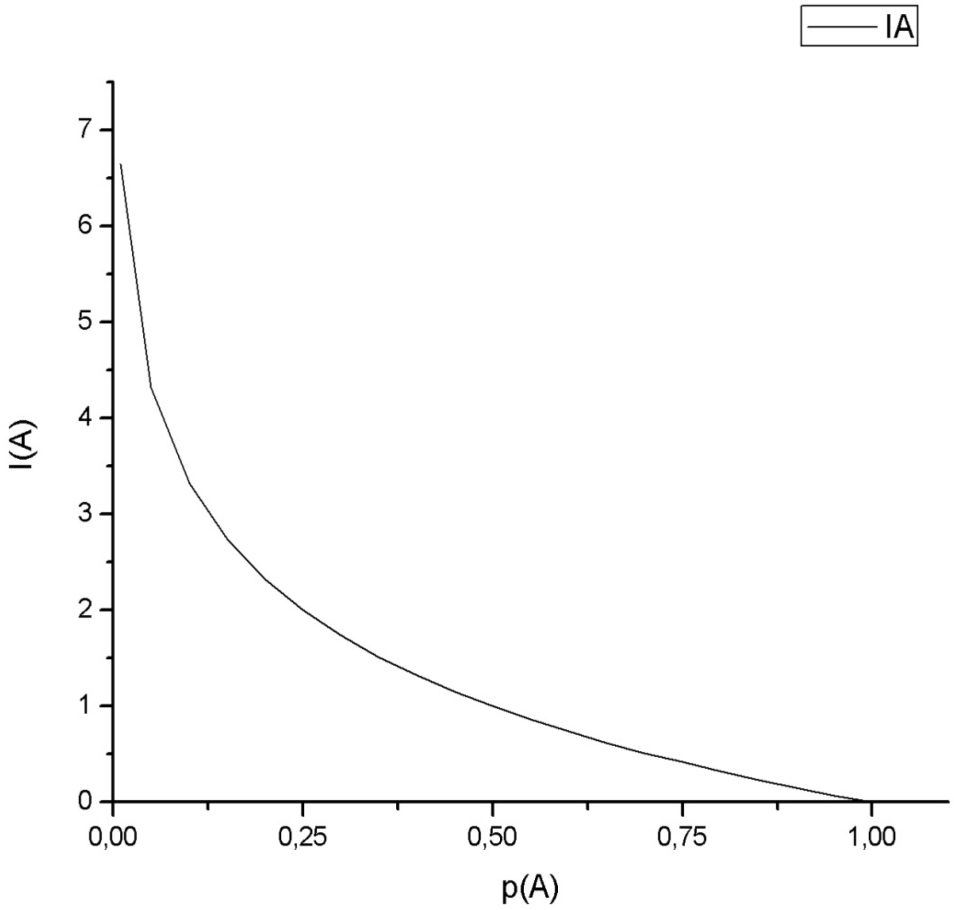
Μέτρο Πληροφορίας

- Η πληροφορία I_A την οποία αποκτούμε από την πραγματοποίηση ενός γεγονότος Α το οποίο έχει πιθανότητα P_A δίνεται από τον τύπο

$$I_A = \log_b\left(\frac{1}{P_A}\right) \equiv -\log_b P_A$$

- Η βάση του λογαρίθμου ορίζει τη μονάδα μέτρησης της πληροφορίας
- Εάν η βάση είναι το 2, τότε η πληροφορία μετριέται σε bits
- Εαν χρησιμοποιηθούν φυσικοί λογάριθμοι (*ln*), τότε η πληροφορία μετριέται σε nat. *(1 nat = 1,44 bits)*
- Στη συνέχεια όλοι οι λογάριθμοι θα είναι με βάση το 2

Γραφική παράσταση της συνάρτησης του μέτρου της πληροφορίας για b = 2



Μέση πληροφορία ή Εντροπία

- Από ένα τηλεπικοινωνιακό σύστημα μεταδίδονται σειρές συμβόλων
- Για μια πηγή πληροφορίας με αλφάβητο

$$A = \{\alpha_1, \alpha_2, \ldots., \alpha_N\}$$

- Η εντροπία ή μέση πληροφορία του A, είναι ο μέσος όρος της αυτοπληροφορίας των συμβόλων στην έξοδο και ορίζεται ως

$$\mathrm{H}(\mathrm{A}) = \mathrm{H}(\alpha_1, \alpha_2, \ldots., \alpha_\mathrm{N}) = -\sum_{n=1}^{N} p(a_n) \log p(a_n)$$

Βασικές Ιδιότητες Εντροπίας

- Η εντροπία είναι το μέτρο της μέσης αβεβαιότητας μιας τυχαίας μεταβλητής
- Ορίζει τον μέσο αριθμό bits που απαιτούνται για να περιγράψουν την τυχαία μεταβλητή
- Η εντροπία μιας τυχαίας μεταβλητής εξαρτάται μόνο από τις πιθανότητες που χαρακτηρίζουν την τυχαία μεταβλητή
  - και όχι από τις πιθανές τιμές της τυχαίας μεταβλητής
- Αφού $\lim_{x \to 0}(x \log x) \to 0$ δεν αλλάζουν την εντροπία οι όροι με μηδενική πιθανότητα.

Δυαδική πηγή Shannon

- Εμφανίζονται δύο έξοδοι με πιθανότητες

  p και $(1 - p)$ αντίστοιχα.

- Η εντροπία είναι

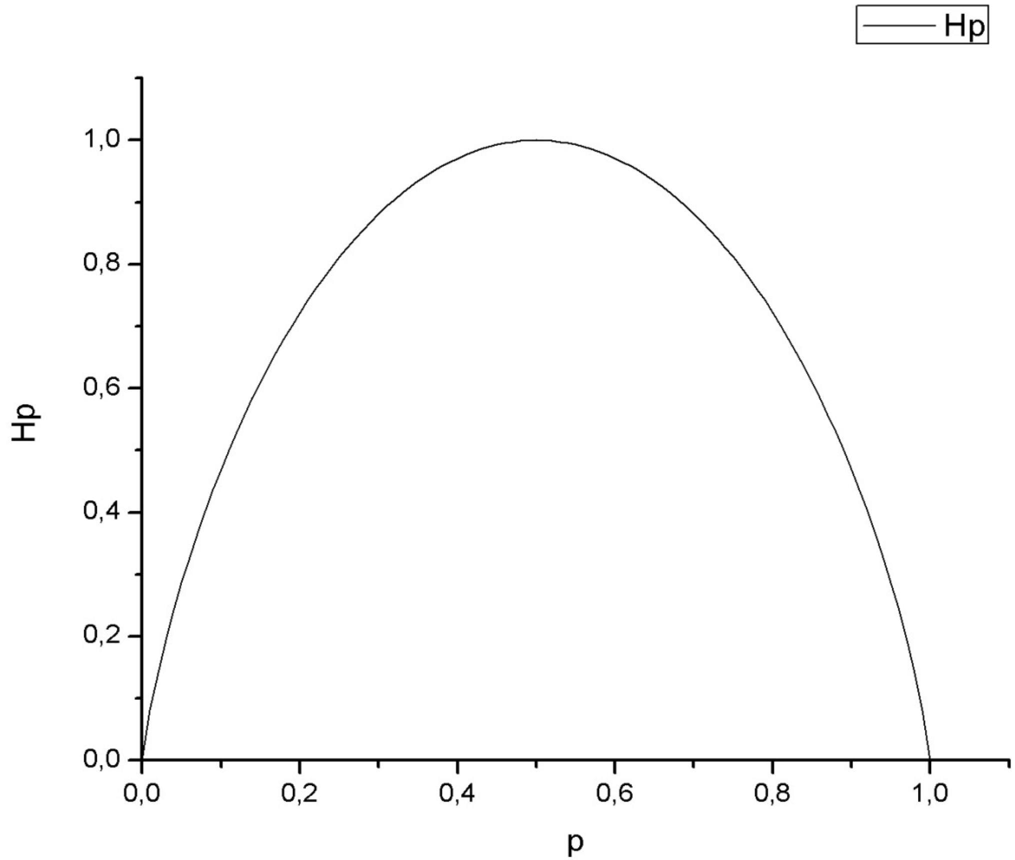
$$H(X) = -p\log(p) - (1 - p)\log(1 - p) = H_b(p)$$

- Μεγιστοποιείται όταν $p = ½$, δηλαδή

$$H_b(1/2) = 1$$

Όπως αναμένεται το αποτέλεσμα μπορεί να μεταφερθεί με 1 bit

Γραφική Παράσταση Συνάρτησης Shannon



Ιδιότητες συνάρτησης Shannon

1. Όταν $p = 0$, είναι $H(X) = 0$. Αυτό προκύπτει από τη σχέση $\lim_{x \to 0} x \log_2 x = 0$
2. Όταν $p = 1$, είναι $H(X) = 0$. Η μεταβλητή παύει να είναι τυχαία και η αβεβαιότητα μηδενίζεται
3. Η εντροπία $H(X)$ λαμβάνει τη μέγιστη τιμή της $H_{max}=1$ bit, όταν $p_o=p_1=1/2$ δηλαδή όταν τα σύμβολα 1 και 0 είναι ισοπίθανα
4. Υπάρχει συμμετρία γύρω από το $p = 0.5$
5. Είναι μία κοίλη συνάρτηση της πιθανότητας

Ιδιότητες της μέσης ποσότητας πληροφορίας ή εντροπίας

- Η μέση πληροφορία H(x) είναι συνεχής συνάρτηση των πιθανοτήτων $p(x_n)$, όπου $n = 1, 2, \ldots.., N$
- Η μέση πληροφορία H(x) είναι συμμετρική συνάρτηση των πιθανοτήτων
  - διαφορετικές τυχαίες μεταβλητές με κατανομές πιθανοτήτων που προέρχονται από μεταθέσεις της ίδιας κατανομής πιθανοτήτων έχουν ίση εντροπία.
- Η εντροπία H(x) παίρνει τη μέγιστη τιμή, όταν όλα τα ενδεχόμενα είναι ισοπίθανα.
  - Τότε, η αβεβαιότητα είναι η μέγιστη δυνατή και κατά συνέπεια, η επιλογή ενός μηνύματος προσφέρει τη μέγιστη δυνατή μέση πληροφορία.

Άσκηση

- Ποια η τιμή της εντροπίας μια πηγής που εκπέμπει 4 σύμβολα Α, Β, Γ & Δ και πιθανότητες εμφάνισης αυτών 1/2 , 1/4, 1/8 & 1/8 αντίστοιχα;

- Πότε η εντροπία της συγκεκριμένης πηγής γίνεται μέγιστη.

Ιδιότητες της μέσης ποσότητας πληροφορίας ή εντροπίας

- ισχύει η αρχή της προσθετικότητας

$$H(X, Y) = H(X) + H(Y)$$

- εάν έχουμε δύο ανεξάρτητα γεγονότα με πιθανότητα p_x και p_y,
- τότε η συνολική πληροφορία που μας δίνουν τα δύο γεγονότα είναι το άθροισμα των επιμέρους πληροφοριών.

Από κοινού εντροπία

- Η από κοινού εντροπία δύο τυχαίων μεταβλητών είναι

$$\mathrm{H}(\mathrm{X},\mathrm{Y}) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p(x_i, y_j)\log(p(x_i, y_j))$$

- Στη γενική περίπτωση Κ τυχαίων μεταβλητών

$$H(X_1 X_2 ....X_K) = -\sum_{i_1=1}^{N_1}\sum_{i_2=2}^{N_2}.....\sum_{i_K=1}^{N_1} p(x_{i1}, x_{i2}, ...x_{iK})\log(p(x_{i1}, x_{i2}, ..., x_{iK}))$$

Υπό Συνθήκη Εντροπία

Η υπό συνθήκη εντροπία της σύνθετης πηγής όταν είναι γνωστή η έξοδος της απλής πηγής δίνεται από την παρακάτω σχέση

$$H(Y\,|\,X) = -\sum_{i=1}^{N} H(Y\,|\,x_i)p(x_i) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p(x_i, y_j)\log(p(y_j / x_i))$$

Κανόνας της Αλυσίδας

Η από κοινού και υπό συνθήκη εντροπία συνδέονται μεταξύ τους.

Το θεώρημα που μας δίνει αυτή τη σύνδεση λέγεται κανόνας της αλυσίδας

δίνεται από την παρακάτω σχέση

$$H(X/Y) = H(X,Y) - H(Y)$$

Αμοιβαία Πληροφορία

- Δίνει την ποσότητα πληροφορίας που μια τυχαία μεταβλητή παρέχει για μια άλλη.
- Μέτρο εξάρτησης των δύο μεταβλητών

$$I(X;Y) = H(X) - H(X|Y)$$

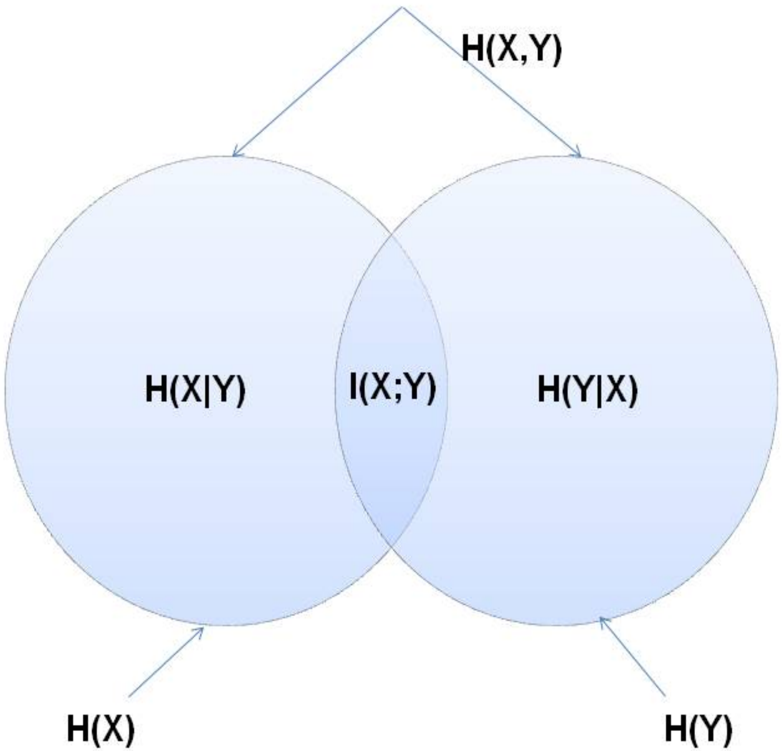
- Είναι συμμετρική ως προς X και Y, δηλαδή

$$I(X;Y) = I(Y;X)$$

- Είναι μη αρνητική, δηλαδή

$$I(X;Y) \geq 0$$

Διάγραμμα Venn



Διάγραμμα Venn

1. Τη **σύνθετη εντροπία (joint entropy) H(X,Y)**, η οποία μετράει τη συνολική πληροφορία των X και Y.

1. Την **εντροπία υπό συνθήκης (conditional entropy) H(X|Y)**, η οποία μετράει την πληροφορία του X, όταν η Y είναι γνωστή και αντιστρόφως.

2. Την **αμοιβαία-κοινή εντροπία (mutual entropy) I(X;Y)**, η οποία μετράει τη σχέση των X και Y, υπό την έννοια ότι μας δείχνει πόσο μειώνεται η πληροφορία του X όταν μαθαίνουμε το Y και αντιστρόφως.