ΠΑΝΕΠΙΣΤΉΜΙΟ ΠΕΛΟΠΟΝΝΉΣΟΥ

Εντροπία και Αμοιβαία Πληροφορία

Θεωρία Πληροφορίας και Κωδίκων

Εαρινό Εξάμηνο

Τμήμα Πληροφορικής και Τηλεπικοινωνιών

Νικόλαος Χ. Σαγιάς
Καθηγητής
Webpage: https://eclass.uop.gr/courses/DIT221/
e-mail: nsagias@uop.gr

27/4/2020 11:36:38 πμ

Από Κοινού Εντροπία

- Η από κοινού μελέτη τυχαίων μεταβλητών (ΤΜ) είναι χρήσιμη στα τηλεπικοινωνιακά συστήματα
- Με τα σύμβολα X και Y μοντελοποιούμε τη λειτουργία της πηγής πληροφορίας και του δέκτη, αντίστοιχα
- Ο δέκτης γνωρίζει κάθε φορά μια τιμή της ΤΜ Y και αποφασίζει υπέρ του συμβόλου που παρήγαγε η πηγή πληροφορίας
- Το κανάλι επικοινωνίας «διαμορφώνει» το βαθμό συσχέτισης μεταξύ των ΤΜ X και Y

Από Κοινού Εντροπία

- Έστω πηγή που παράγει n σύμβολα $\mathscr{L}_X = \{x_1, x_2, \ldots, x_n\}$ με πιθανότητες $\mathbb{P}\{X = x_i\} = p(x_i)$, με $p(x_i)$ την περιθώρια συνάρτηση πιθανότητας μάζας (ΣΠΜ) της τυχαίας μεταβλητής (ΤΜ) X

- Έστω δεύτερη πηγή που παράγει m σύμβολα $\mathscr{L}_y = \{y_1, y_2, \ldots, y_m\}$ με πιθανότητες εμφάνισης $\mathbb{P}\{Y = y_j\} = p(y_j)$, με $p(y_j)$ την περιθώρια ΣΠΜ της ΤΜ Y

- Μπορούμε να ορίσουμε τον δισδιάστατο δειγματοχώρο των κοινών γεγονότων X και Y ως

$$\mathscr{L}_{X,Y} = \{(x_1,y_1), (x_1,y_2), \ldots, (x_1,y_m), (x_2,y_1), (x_2,y_2), \ldots, (x_2,y_m), \ldots, (x_n,y_1), (x_n,y_2), \ldots, (x_n,y_m)\}$$

- Η από κοινού ΣΠΜ των X και Y είναι η πιθανότητα

$$\mathbb{P}\{X = x_i, Y = y_j\} = p(x_i, y_j)$$

Από Κοινού Εντροπία

$$\sum_{j=1}^{m} p(x_i, y_j) = p(x_i)$$

X \ Y	y_1	y_2	–	y_m	
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	–	$p(x_1, y_m)$	$p(x_1)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	–	$p(x_2, y_m)$	$p(x_2)$
|	|	|		|	|
x_n	$p(x_n, y_1)$	$p(x_n, y_2)$	–	$p(x_n, y_m)$	$p(x_n)$
	$p(y_1)$	$p(y_2)$		$p(y_m)$	1

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

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

Από Κοινού Εντροπία

- Προφανώς, η από κοινού ΣΠΜ των X και Y έχει την εξής ιδιότητα

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p\left(x_i, y_j\right) = 1$$

- Κάποιες επιπλέον ιδιότητες για την $p(x_i, y_j)$ και τις περιθώριες ΣΠΜ της είναι οι εξής:
  - Αν αθροίσουμε σε όλα τα ενδεχόμενα x_i προκύπτει

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

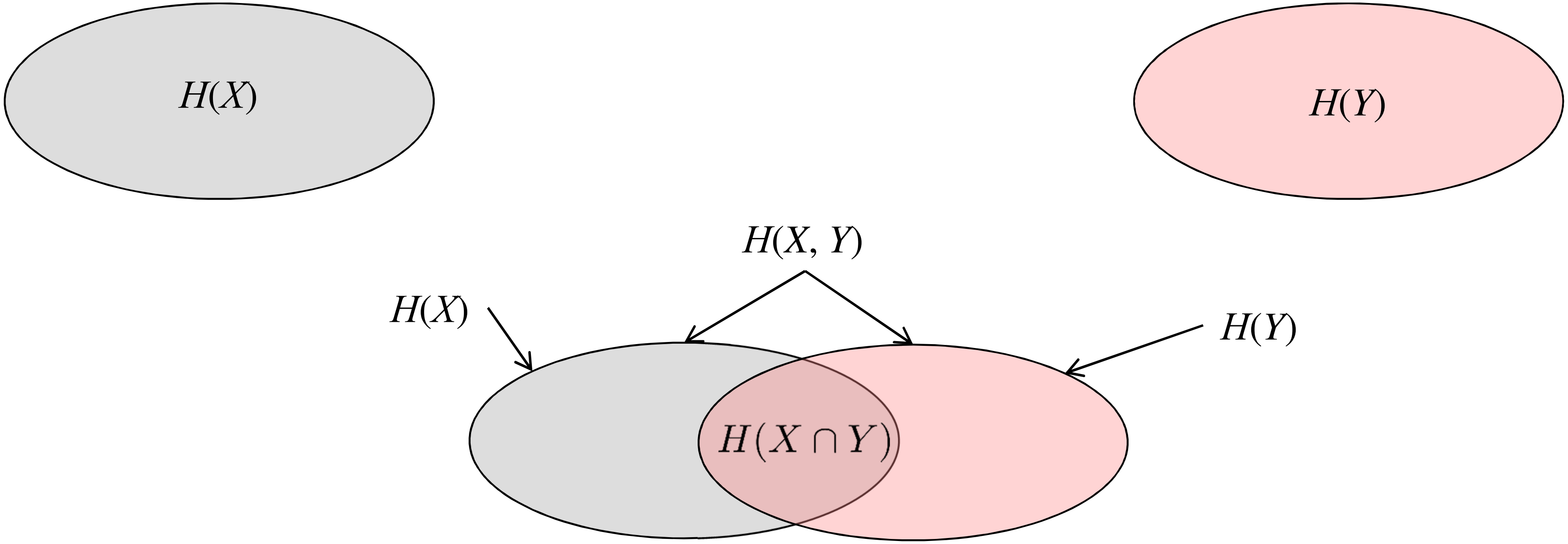
  - Αν αθροίσουμε σε όλα τα ενδεχόμενα y_j προκύπτει

$$\sum_{j=1}^{m} p\left(x_i, y_j\right) = p\left(x_i\right)$$

Από Κοινού Εντροπία

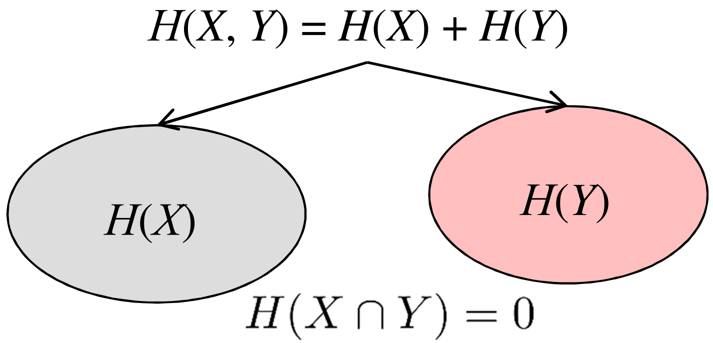
- Η από κοινού εντροπία (*joint entropy*) του ζεύγους ΤΜ *X* και *Y* ορίζεται ως

$$H(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p\left(x_i, y_j\right) \log_2\left[\frac{1}{p\left(x_i, y_j\right)}\right]$$



Από Κοινού Εντροπία

- Έστω το ζεύγος των ανεξάρτητων ΤΜ X και Y



- Η από κοινού ΣΠΜ γράφεται

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

- Στην περίπτωση αυτή, η από κοινού εντροπία είναι

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\, \log_2 [p(x_i, y_j)] = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)\, p(y_j)\, \log_2 [p(x_i)\, p(y_j)] =$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)\, p(y_j)\, \log_2 [p(x_i)] - \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)\, p(y_j)\, \log_2 [p(y_j)] =$$

$$= -\sum_{i=1}^{n} p(x_i)\, \log_2 [p(x_i)] - \sum_{j=1}^{m} p(y_j)\, \log_2 [p(y_j)] = H(X) + H(Y)$$

Από Κοινού Εντροπία

- <u>Παράδειγμα:</u> Έστω το αποτέλεσμα ρίψης 2 ζαριών (ανεξάρτητα γεγονότα) με δειγματοχώρους

$$\Omega_X = \{1, 2, \ldots, 6\} \quad \text{και} \quad \Omega_Y = \{1, 2, \ldots, 6\}$$

- Για κάθε ζάρι, όλα τα αποτελέσματα εμφανίζονται με ίδια πιθανότητα $p(x_i) = p(y_j) = P = 1/6$
- Η εντροπία για κάθε ζάρι είναι

$$H(X) = H(Y) = -\sum_{i=1}^{6} P \log_2(P) = -\sum_{i=1}^{6} \frac{1}{6} \log_2\left(\frac{1}{6}\right) \approx 2.585 \text{ bit}$$

- Όλα τα κοινά αποτελέσματα από τα 2 ζάρια έχουν από κοινού ΣΠΜ

$$p(x_i, y_j) = p(x_i)\, p(y_j) = 1/6 \times 1/6 = 1/36$$

- Συνεπώς, η από κοινού εντροπία είναι

$$H(X,Y) = -\sum_{i=1}^{6}\sum_{j=1}^{6} p(x_i, y_j) \log_2[p(x_i, y_j)] = -\sum_{i=1}^{6}\sum_{j=1}^{6} \frac{1}{36} \log_2\left(\frac{1}{36}\right) \approx 5.170 \text{ bit}$$

- Δηλαδή, όντως, η από κοινού εντροπία είναι ίση με το άθροισμα των επιμέρους εντροπιών

$$H(X, Y) = H(X) + H(Y) \approx 5.17 \text{ bit}$$

Από Κοινού Εντροπία

- <u>Παράδειγμα:</u> Έστω το ζεύγος των εξαρτημένων ΤΜ X και Y με την από κοινού ΣΠΜ $p(x_i, y_j)$ του πίνακα

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	
	y_2	1/16	1/8	1/32	1/32	
	y_3	1/16	1/16	1/16	1/16	
	y_4	1/4	0	0	0	
	$p(x_i)$					1

- Αθροίζοντας σε όλα τα στοιχεία της στήλης i, βρίσκουμε την πιθανότητα εμφάνισης του x_i
- Αθροίζοντας σε όλα τα στοιχεία της γραμμής j, βρίσκουμε την πιθανότητα εμφάνισης του y_j

Από Κοινού Εντροπία

- <u>Παράδειγμα (*συνέχεια*):</u>

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	
	y_2	1/16	1/8	1/32	1/32	
	y_3	1/16	1/16	1/16	1/16	
	y_4	1/4	0	0	0	
	$p(x_i)$	1/2				1

- Αθροίζοντας σε όλα τα στοιχεία της 1ης στήλης, βρίσκουμε την πιθανότητα εμφάνισης του x_1

Από Κοινού Εντροπία

- <u>Παράδειγμα (*συνέχεια*):</u>

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	
	y_2	1/16	1/8	1/32	1/32	
	y_3	1/16	1/16	1/16	1/16	
	y_4	1/4	0	0	0	
	$p(x_i)$	1/2	1/4			1

- Αθροίζοντας σε όλα τα στοιχεία της 2ης στήλης, βρίσκουμε την πιθανότητα εμφάνισης του x_2

Από Κοινού Εντροπία

- <u>Παράδειγμα (*συνέχεια*):</u>

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	
	y_2	1/16	1/8	1/32	1/32	
	y_3	1/16	1/16	1/16	1/16	
	y_4	1/4	0	0	0	
	$p(x_i)$	1/2	1/4	1/8		1

- Αθροίζοντας σε όλα τα στοιχεία της 3ης στήλης, βρίσκουμε την πιθανότητα εμφάνισης του x_3

Από Κοινού Εντροπία

- <u>Παράδειγμα (*συνέχεια*):</u>

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	
	y_2	1/16	1/8	1/32	1/32	
	y_3	1/16	1/16	1/16	1/16	
	y_4	1/4	0	0	0	
	$p(x_i)$	1/2	1/4	1/8	1/8	1

$\underbrace{1/2 \quad 1/4 \quad 1/8 \quad 1/8}_{1}$

- Αθροίζοντας σε όλα τα στοιχεία της 4ης στήλης, βρίσκουμε την πιθανότητα εμφάνισης του x_4

Από Κοινού Εντροπία

- <u>Παράδειγμα (*συνέχεια*):</u>

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	1/4
	y_2	1/16	1/8	1/32	1/32	
	y_3	1/16	1/16	1/16	1/16	
	y_4	1/4	0	0	0	
	$p(x_i)$	1/2	1/4	1/8	1/8	1

- Αθροίζοντας σε όλα τα στοιχεία της 1ης γραμμής, βρίσκουμε την πιθανότητα εμφάνισης του y_1

Από Κοινού Εντροπία

- <u>Παράδειγμα (*συνέχεια*):</u>

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	1/4
	y_2	1/16	1/8	1/32	1/32	1/4
	y_3	1/16	1/16	1/16	1/16	
	y_4	1/4	0	0	0	
	$p(x_i)$	1/2	1/4	1/8	1/8	1

- Αθροίζοντας σε όλα τα στοιχεία της 2^{ης} γραμμής, βρίσκουμε την πιθανότητα εμφάνισης του y_2

Από Κοινού Εντροπία

- <u>Παράδειγμα (*συνέχεια*):</u>

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	1/4
	y_2	1/16	1/8	1/32	1/32	1/4
	y_3	1/16	1/16	1/16	1/16	1/4
	y_4	1/4	0	0	0	
	$p(x_i)$	1/2	1/4	1/8	1/8	1

- Αθροίζοντας σε όλα τα στοιχεία της 3ης γραμμής, βρίσκουμε την πιθανότητα εμφάνισης του y_3

Από Κοινού Εντροπία

- <u>Παράδειγμα (*συνέχεια*):</u>

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	1/4
	y_2	1/16	1/8	1/32	1/32	1/4
	y_3	1/16	1/16	1/16	1/16	1/4
	y_4	1/4	0	0	0	1/4
	$p(x_i)$	1/2	1/4	1/8	1/8	1

1

- Αθροίζοντας σε όλα τα στοιχεία της 4^{ης} γραμμής, βρίσκουμε την πιθανότητα εμφάνισης του y_4

Από Κοινού Εντροπία

- <u>Παράδειγμα (*συνέχεια*):</u>

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	1/4
	y_2	1/16	1/8	1/32	1/32	1/4
	y_3	1/16	1/16	1/16	1/16	1/4
	y_4	1/4	0	0	0	1/4
	$p(x_i)$	1/2	1/4	1/8	1/8	1

- Η από κοινού εντροπία με $n = m = 4$ είναι

$$
\begin{aligned}
H(X,Y) = -\sum_{i=1}^{4}\sum_{j=1}^{4} p\left(x_i,y_j\right)\,\log_2\left[p\left(x_i,y_j\right)\right] = \\
= -\frac{1}{8}\,\log_2\left(\frac{1}{8}\right) - \frac{1}{16}\,\log_2\left(\frac{1}{16}\right) - \frac{1}{32}\,\log_2\left(\frac{1}{32}\right) - \frac{1}{32}\,\log_2\left(\frac{1}{32}\right) \\
-\frac{1}{16}\,\log_2\left(\frac{1}{16}\right) - \frac{1}{8}\,\log_2\left(\frac{1}{8}\right) - \frac{1}{32}\,\log_2\left(\frac{1}{32}\right) - \frac{1}{32}\,\log_2\left(\frac{1}{32}\right) \\
-\frac{1}{16}\,\log_2\left(\frac{1}{16}\right) - \frac{1}{16}\,\log_2\left(\frac{1}{16}\right) - \frac{1}{16}\,\log_2\left(\frac{1}{16}\right) - \frac{1}{16}\,\log_2\left(\frac{1}{16}\right) \\
-\frac{1}{4}\,\log_2\left(\frac{1}{4}\right) = \frac{27}{8}\ \text{bit}
\end{aligned}
$$

Από Κοινού Εντροπία

- <u>Παράδειγμα (*συνέχεια*):</u> Προηγουμένως βρήκαμε ότι

$$H(X, Y) = 27/8 \text{ bit}$$

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	1/4
	y_2	1/16	1/8	1/32	1/32	1/4
	y_3	1/16	1/16	1/16	1/16	1/4
	y_4	1/4	0	0	0	1/4
	$p(x_i)$	1/2	1/4	1/8	1/8	1

- Η εντροπία της ΤΜ *X* είναι

$$H(X) = -\sum_{i=1}^{4} p(x_i) \log_2 [p(x_i)] = -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{4} \log_2 \left(\frac{1}{4}\right) - \frac{1}{8} \log_2 \left(\frac{1}{8}\right) - \frac{1}{8} \log_2 \left(\frac{1}{8}\right) = \frac{7}{4} \text{ bit}$$

- Η εντροπία της ΤΜ *Y* είναι

$$H(Y) = -\sum_{j=1}^{4} p(y_j) \log_2 [p(y_j)] = -\frac{1}{4} \log_2 \left(\frac{1}{4}\right) - \frac{1}{4} \log_2 \left(\frac{1}{4}\right) - \frac{1}{4} \log_2 \left(\frac{1}{4}\right) - \frac{1}{4} \log_2 \left(\frac{1}{4}\right) = 2 \text{ bit}$$

- Παρατηρούμε ότι

$$H(X,Y) < H(X) + H(Y) = 30/8 \text{ bit}$$

- Συνεπώς, η εξάρτηση μεταξύ γεγονότων μειώνει την αβεβαιότητα και άρα την από κοινού εντροπία

Υπό Συνθήκη Εντροπία

- Η ποσότητα πληροφορίας του αποτελέσματος, x_i με δεδομένο το αποτέλεσμα y_j είναι

$$I(x_i|y_j) = -\log_2[p(x_i|y_j)]$$

- Η εντροπία της ΤΜ X, δεδομένου του αποτελέσματος y_j είναι

$$H\left(X|y_j\right) = \sum_{i=1}^{n} p(x_i|y_j)\, I(x_i|y_j)$$

- Λαμβάνοντας υπόψιν όλα τα δεδομένα αποτελέσματα της ΤΜ Y, η εντροπία της ΤΜ X δεδομένης της ΤΜ Y είναι

$$H(X|Y) = -\sum_{j=1}^{m} p\left(y_j\right) \sum_{i=1}^{n} p\left(x_i|y_j\right)\, \log_2\left[p\left(x_i|, y_j\right)\right]$$

Υπό Συνθήκη Εντροπία

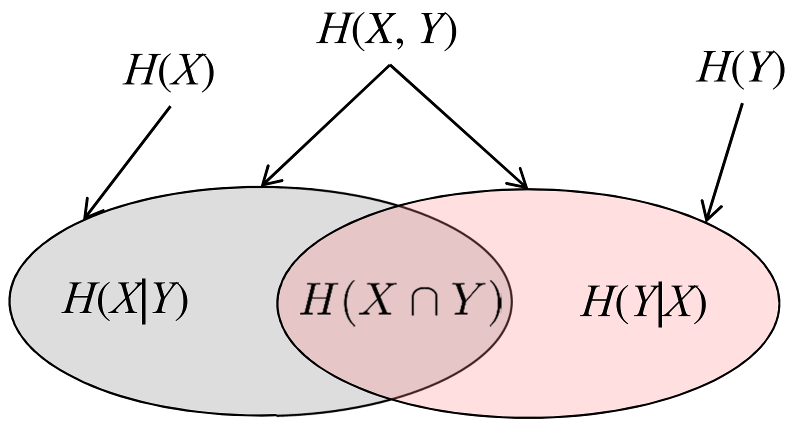
- Δεδομένου ότι $p(x_i, y_j) = p(x_i \mid y_j)\, p(y_j)$, η προηγούμενη σχέση για την υπό συνθήκη εντροπία (*conditional entropy*) μπορεί επίσης να εκφραστεί και ως

$$H(X|Y) = -\sum_{j=1}^{m}\sum_{i=1}^{n} p(x_i, y_j)\, \log_2 [p(x_i|y_j)]$$

- Η υπό συνθήκη εντροπία εκφράζει το πόσο αβέβαιοι είμαστε για την ΤΜ *X*, όταν γνωρίζουμε την *Y*

- Γενικά

$$H(X|Y) \neq H(Y|X)$$



Υπό Συνθήκη Εντροπία

- <u>Παράδειγμα:</u> Η από κοινού εντροπία βρέθηκε $H(X, Y) = 27/8$ bit
- Η υπό συνθήκη εντροπία $H(X|Y)$ μπορεί να βρεθεί αν πρώτα υπολογιστεί η περιθώρια ΣΠΜ $p(y_j)$

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

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	1/4
	y_2	1/16	1/8	1/32	1/32	1/4
	y_3	1/16	1/16	1/16	1/16	1/4
	y_4	1/4	0	0	0	1/4
	$p(x_i)$	1/2	1/4	1/8	1/8	1

- Η υπό συνθήκη εντροπία $H(X|Y)$ μπορεί να υπολογιστεί όπως παρακάτω

$$H(X|Y) = -\sum_{j=1}^{m}\sum_{i=1}^{n} p(x_i, y_j)\,\log_2[p(x_i|, y_j)] = -\sum_{j=1}^{m}\sum_{i=1}^{n} p(x_i, y_j)\,\log_2\left[\frac{p(x_i, y_j)}{p(y_j)}\right]$$

$$= -\frac{1}{8}\log_2\left(\frac{1/8}{1/4}\right) - \frac{1}{16}\log_2\left(\frac{1/16}{1/4}\right) - \frac{1}{32}\log_2\left(\frac{1/32}{1/4}\right) - \frac{1}{32}\log_2\left(\frac{1/32}{1/4}\right)$$

$$-\frac{1}{16}\log_2\left(\frac{1/16}{1/4}\right) - \frac{1}{8}\log_2\left(\frac{1/8}{1/4}\right) - \frac{1}{32}\log_2\left(\frac{1/32}{1/4}\right) - \frac{1}{32}\log_2\left(\frac{1/32}{1/4}\right)$$

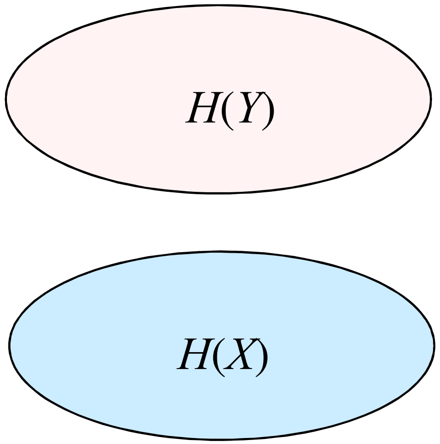
$$-\frac{1}{16}\log_2\left(\frac{1/16}{1/4}\right) - \frac{1}{16}\log_2\left(\frac{1/16}{1/4}\right) - \frac{1}{16}\log_2\left(\frac{1/16}{1/4}\right) - \frac{1}{16}\log_2\left(\frac{1/16}{1/4}\right) - \frac{1}{4}\log_2\left(\frac{1/4}{1/4}\right) = \frac{11}{8}\text{ bit} \le H(X) = \frac{14}{8}\text{ bit}$$

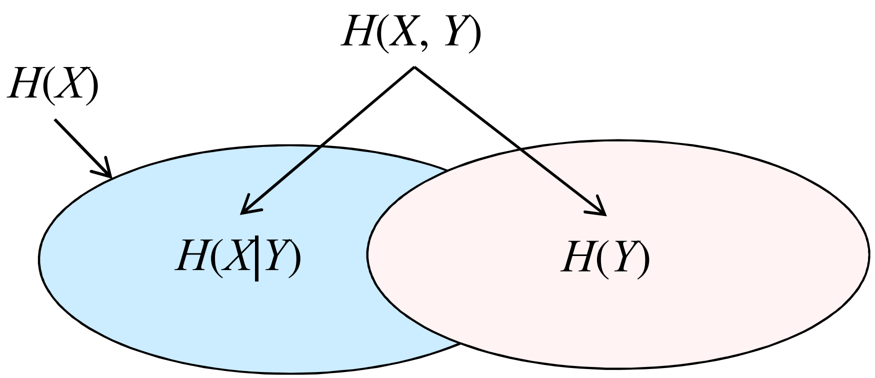
Από Κοινού Εντροπία

- Κανόνας Αλυσίδας: Η από κοινού εντροπία του ζεύγους ΤΜ X και Y μπορεί να εκφραστεί ως ακολούθως

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

- Γραφικά, το 1[ο] ίσον της παραπάνω ιδιότητα αποτυπώνεται στο διάγραμμα Venn





Από Κοινού Εντροπία

- Απόδειξη: Εξ' ορισμού, η από κοινού εντροπία του ζεύγους γεγονότων X και Y είναι

$$
\begin{aligned}
H(X,Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\,\log_2\left[p(x_i,y_j)\right] \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\,\log_2\left[p(x_i|y_j)\;p(y_j)\right] \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\,\log_2\left[p(y_j)\right] - \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\,\log_2\left[p(x_i|y_j)\right] \\
&= \underbrace{-\sum_{j=1}^{m}\log_2\left[p(y_j)\right]\overbrace{\sum_{i=1}^{n} p(x_i,y_j)}^{p(y_j)}}_{H(Y)}\;\underbrace{-\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\,\log_2\left[p(x_i|y_j)\right]}_{H(X/Y)}
\end{aligned}
$$

- Παρομοίως αποδεικνύεται επιπλέον ότι $H(X,Y) = H(X) + H(Y|X)$

Υπό Συνθήκη Εντροπία

- Η συνάρτηση $f(x)$ χαρακτηρίζεται ως κοίλη (*concave*) όταν για κάθε x_1, x_2 ισχύει η ανισότητα

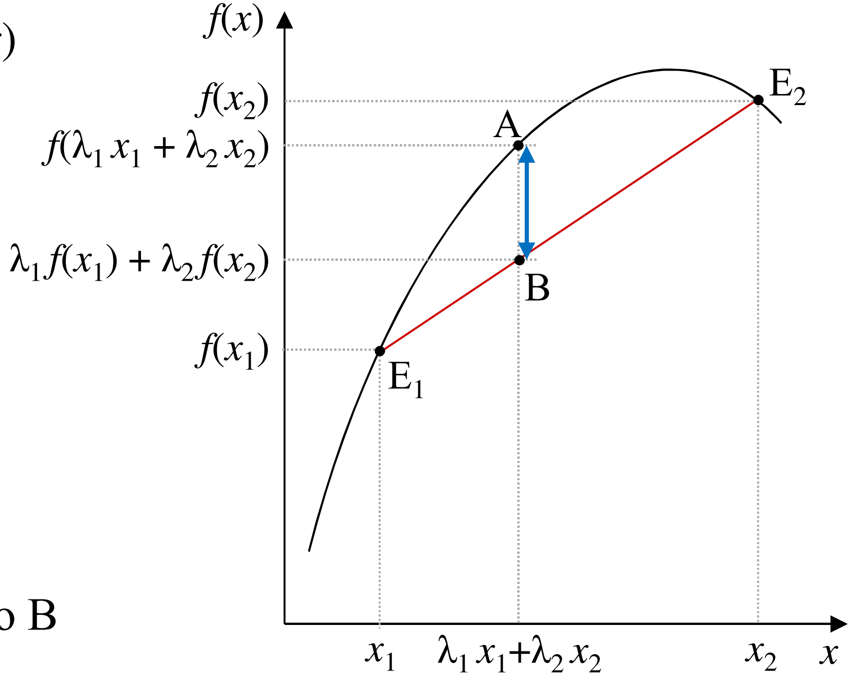
$$f(\lambda_1 x_1 + \lambda_2 x_2) \geq \lambda_1 f(x_1) + \lambda_2 f(x_2)$$

  με $\lambda_1 + \lambda_2 = 1$ και $0 \leq \lambda_1, \lambda_2 \leq 1$

- Πρακτικά, σημαίνει ότι οποιοδήποτε ευθύγραμμο τμήμα E_1E_2 μεταξύ 2 σημείων της $f(x)$ θα βρίσκεται κάτω από την $f(x)$

- Συγκεκριμένα:
  - Η τιμή $\lambda_1 x_1 + \lambda_2 x_2$ βρίσκεται μεταξύ x_1 και x_2
  - Το $A(\lambda_1 x_1 + \lambda_2 x_2, f(\lambda_1 x_1 + \lambda_2 x_2))$ είναι σημείο της $f(x)$
  - Το $B(\lambda_1 x_1 + \lambda_2 x_2, \lambda_1 f(x_1) + \lambda_2 f(x_2))$ είναι σημείο του E_1E_2

- Βάσει της ανισότητας το σημείο A βρίσκεται πάνω από το B



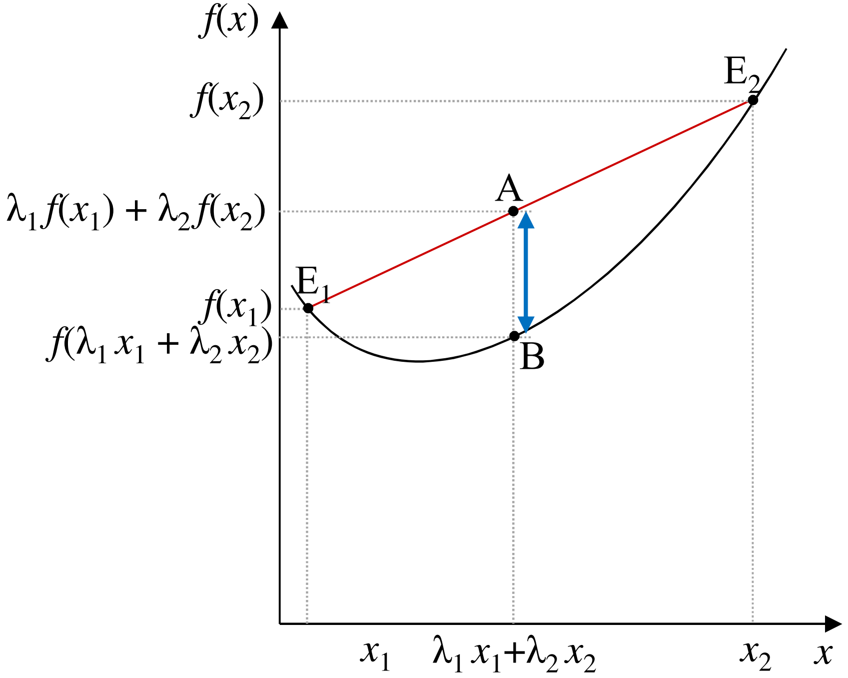
Υπό Συνθήκη Εντροπία

- Η συνάρτηση $f(x)$ χαρακτηρίζεται ως κυρτή (*convex*) όταν για κάθε x_1, x_2 ισχύει η ανισότητα

$$f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2)$$

με $\lambda_1 + \lambda_2 = 1$ και $0 \leq \lambda_1, \lambda_2 \leq 1$ (δηλαδή το ευθύγραμμο τμήμα E_1E_2 βρίσκεται πάνω από την $f(x)$)

- Ο έλεγχος περί κοίλων ή κυρτών συναρτήσεων γίνεται βάσει του πρόσημου της 2ης παραγώγου



- Παράδειγμα #1: Η 2η παράγωγος της $f(x) = x^2$,

είναι $f''(x) = 2 > 0$, άρα η $f(x)$ είναι κυρτή

- Παράδειγμα #2: Η 2η παράγωγος της $f(x) = \log_2(x)$,

είναι $f''(x) = -\log_2(e)/x^2 < 0$, άρα η $f(x)$ είναι κοίλη

Υπό Συνθήκη Εντροπία

- Γενικεύοντας, ισχύει η ανισότητα Jensen για κοίλες συναρτήσεις

$$\sum_{i=1}^{n} \lambda_i\, f(x_i) \leq f\left(\sum_{i=1}^{n} \lambda_i\, x_i\right)$$

με $\sum_{i=1}^{n} \lambda_i = 1$ και $\lambda_i \in [0,1]$ για κάθε $i = 1, 2, \ldots, n$

- Η αντίθετη φορά της ανισότητας ισχύει για κυρτές συναρτήσεις

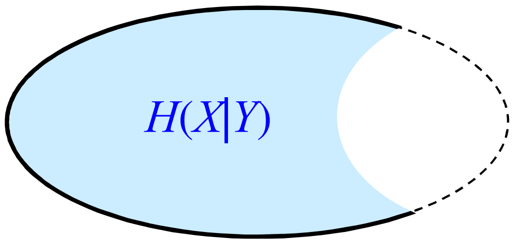
- Η απόδειξη της ανισότητας Jensen μπορεί να γίνει επαγωγικά

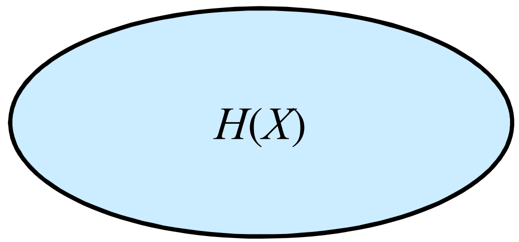
Υπό Συνθήκη Εντροπία

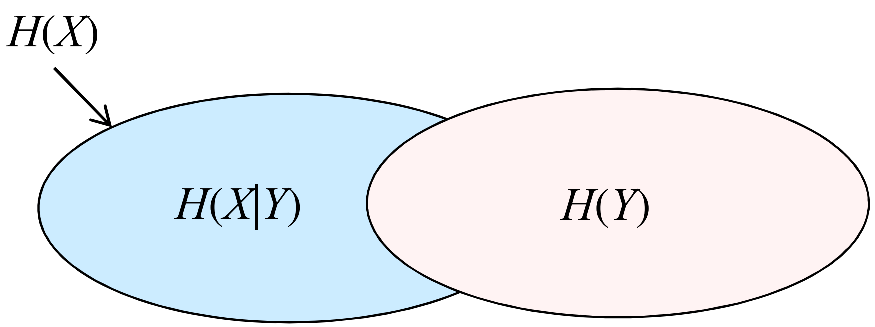
- Για την υπό συνθήκη εντροπία ισχύει η εξής ανισότητα

$$H(X|Y) \leq H(X)$$

- Η ανισότητα αυτή ουσιαστικά ερμηνεύεται ότι ως εξής: Η γνώση μιας συσχετισμένης ΤΜ Y μειώνει την αβεβαιότητα και άρα την εντροπία της ΤΜ X







Υπό Συνθήκη Εντροπία

- Απόδειξη: Η συνάρτηση $\log_2(\cdot)$ είναι κοίλη, οπότε μπορούμε να χρησιμοποιήσουμε την ανισότητα Jensen για κοίλες συναρτήσεις

$$\sum_{j=1}^{m} \lambda_j \, \log_2(z_j) \le \log_2\left(\sum_{j=1}^{m} \lambda_j \, z_j\right)$$

με $\sum_{j=1}^{m} \lambda_j = 1$ και $\lambda_j \in [0,1]$ για κάθε $j = 1, 2, \ldots, m$

- Ξεκινώντας από τη σχέση για την υπό συνθήκη εντροπία και μετά από αλγεβρικές πράξεις έχουμε

$$H(X|Y) = \sum_{j=1}^{m}\sum_{i=1}^{n} p(x_i, y_j)\, \log_2\left[\frac{1}{p(x_i|y_j)}\right] = \sum_{i=1}^{n} p(x_i) \sum_{j=1}^{m} \overbrace{\frac{p(x_i, y_j)}{p(x_i)}}^{0 \le \lambda_j \le 1} \log_2 \overbrace{\left[\frac{p(y_j)}{p(x_i, y_j)}\right]}^{z_j} \le$$

$$\le \sum_{i=1}^{n} p(x_i) \log_2\left[\sum_{j=1}^{m} \underbrace{\frac{p(x_i, y_j)}{p(x_i)}}_{\lambda_j} \underbrace{\frac{p(y_j)}{p(x_i, y_j)}}_{z_j}\right] = \sum_{i=1}^{n} p(x_i) \log_2\left[\frac{1}{p(x_i)} \underbrace{\sum_{j=1}^{m} p(y_j)}_{1}\right] = H(X)$$

Από Κοινού Εντροπία

- Φράγμα ανεξαρτησίας της από κοινού εντροπίας: Η από κοινού εντροπία δεν μπορεί να είναι μεγαλύτερη από το άθροισμα των επί μέρους εντροπιών

$$H(X, Y) \leq H(X) + H(Y)$$

- Η ισότητα ισχύει όταν οι X και Y είναι μεταξύ τους ανεξάρτητες ΤΜ

- Απόδειξη: Η από κοινού εντροπία δύο ΤΜ X και Y γράφεται

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

- Με βάση το γεγονός ότι $H(Y|X) \leq H(Y)$, η προηγούμενη εξίσωση γράφεται

$$H(X, Y) = H(X) + H(Y|X) \leq H(X) + H(Y)$$

Από Κοινού Εντροπία

- Έστω n ΤΜ $\{X_1, X_2, \ldots, X_n\}$ με από κοινού ΣΠΜ $p(x_1, x_2, \ldots, x_n)$
- Κανόνας αλυσίδας για την εντροπία: Η από κοινού εντροπία των n ΤΜ γράφεται

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i | X_{i-1}, X_{i-2}, \ldots, X_1)$$

- Απόδειξη: Θα χρησιμοποιήσουμε τον κανόνα της αλυσίδας για την από κοινού πιθανότητα

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i | x_{i-1}, x_{i-2}, \ldots, x_1)$$

- Ξεκινώντας από την από κοινού εντροπία και αντικαθιστώντας την από κοινού ΣΠΜ έχουμε

$$H(X_1, X_2, \ldots, X_n) = -\sum_{x_1, x_2, \ldots, x_n} p(x_1, x_2, \ldots, x_n) \log_2 [p(x_1, x_2, \ldots, x_n)] =$$

$$= -\sum_{x_1, x_2, \ldots, x_n} p(x_1, x_2, \ldots, x_n) \log_2 \left[\prod_{i=1}^{n} p(x_i | x_{i-1}, x_{i-2}, \ldots, x_1)\right] =$$

Από Κοινού Εντροπία

- <u>Απόδειξη (*συνέχεια*)</u>:

$$= -\sum_{x_1,x_2,\ldots,x_n} p\left(x_1,x_2,\ldots,x_n\right)\,\log_2\left[\prod_{i=1}^{n} p\left(x_i|x_{i-1},x_{i-2},\ldots,x_1\right)\right] =$$

$$= -\sum_{x_1,x_2,\ldots,x_n} p\left(x_1,x_2,\ldots,x_n\right)\,\sum_{i=1}^{n}\log_2\left[p\left(x_i|x_{i-1},x_{i-2},\ldots,x_1\right)\right] =$$

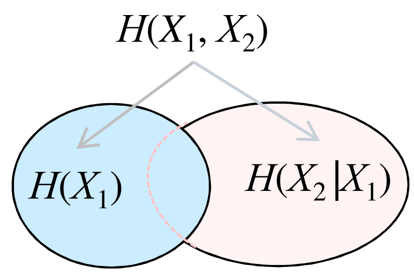
$$= -\sum_{x_1,x_2,\ldots,x_n} p\left(x_1,x_2,\ldots,x_n\right)\,\left\{\log_2\left[p\left(x_1\right)\right]+\log_2\left[p\left(x_2|x_1\right)\right]+\cdots+\log_2\left[p\left(x_n|x_{n-1},x_{n-2},\ldots,x_1\right)\right]\right\} =$$

$$= -\sum_{i=1}^{n}\sum_{x_1,x_2,\ldots,x_i} p\left(x_1,x_2,\ldots,x_i\right)\,\log_2\left[p\left(x_i|x_{i-1},x_{i-2},\ldots,x_1\right)\right] =$$

$$= -\sum_{i=1}^{n} H\left(X_i|X_{i-1},X_{i-2},\ldots,X_1\right)$$

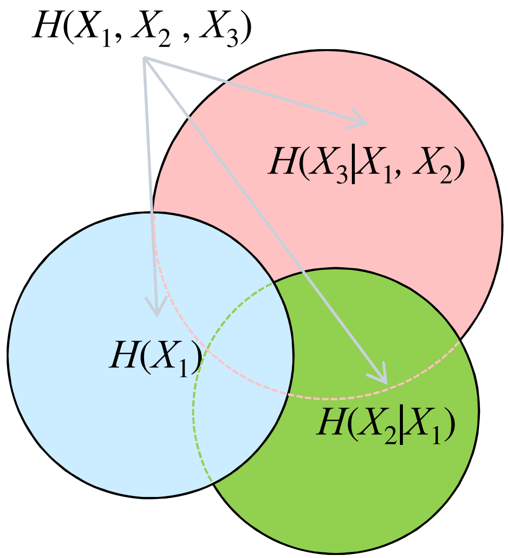
Από Κοινού Εντροπία

- Παραδείγματα εφαρμογής του κανόνα της αλυσίδας



- Για $n = 2$ ΤΜ

$$H\left(X_1, X_2\right) = \sum_{i=1}^{2} H(X_i|X_{i-1}) = H(X_1) + H(X_2|X_1)$$



- Για $n = 3$ ΤΜ

$$H\left(X_1, X_2, X_3\right) = \sum_{i=1}^{3} H(X_i|X_{i-1}, X_{i-2})$$

$$= H(X_1) + H(X_2|X_1) + H(X_3|X_2, X_1)$$

- Για $n = 4$ ΤΜ

$$H\left(X_1, X_2, X_3, X_4\right) = \sum_{i=1}^{4} H(X_i|X_{i-1}, X_{i-2}, X_{i-3}) =$$

$$= H(X_1) + H(X_2|X_1) + H(X_3|X_2, X_1) + H(X_4|X_3, X_2, X_1)$$

Από Κοινού Εντροπία

- Θα γενικεύσουμε την ανισότητα για το φράγμα ανεξαρτησίας της εντροπίας που είχαμε αποδείξει προηγουμένως για δύο ΤΜ $H(X, Y) \leq H(X) + H(Y)$

- Με βάση τον κανόνα αλυσίδας για την εντροπία, η από κοινού εντροπία των n ΤΜ γράφεται

$$H\left(X_1, X_2, \ldots, X_n\right) = \sum_{i=1}^{n} H\left(X_i | X_{i-1}, X_{i-2}, \ldots, X_1\right)$$

- Δεδομένο ότι η υπό συνθήκη εντροπία μιας ΤΜ μειώνεται με τη γνώση άλλων συσχετισμένων ΤΜ ισχύει ότι

$$H(X_i \,|\, X_{i-1}, X_{i-2}, \ldots, X_1) \leq H(X_i)$$

- Επιπλέον, βάσει αναδιάταξης των ΤΜ ισχύει ότι $H(X_1, X_2, \ldots, X_n) \geq H(X_i)$ ($i = 1, 2, \ldots, n$) και άρα

$$H(X_1), H(X_2), \ldots, H(X_n) \leq H(X_1, X_2, \ldots, X_n) \leq H(X_1) + H(X_2) + \cdots + H(X_n)$$

Σχετική Εντροπία

- Έστω κατανομή P με πιθανότητες $\{p_1, p_2, \ldots, p_n\}$ με $\sum_{i=1}^{n} p_i = 1$ και έστω μία δεύτερη κατανομή Q με πιθανότητες $\{q_1, q_2, \ldots, q_n\}$ με $\sum_{i=1}^{n} q_i = 1$

- Η ανισότητα Gibbs δίδεται από την σχέση

$$-\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2(q_i)$$

- <u>Απόδειξη</u>: Κάνοντας χρήση της ανισότητας $\log_2(x) \leq (x-1)\log_2(e)$, με $x = q_i / p_i$, έχουμε

$$\log_2\left(\frac{q_i}{p_i}\right) \leq \left(\frac{q_i}{p_i} - 1\right) \log_2(e) \Leftrightarrow p_i \log_2\left(\frac{q_i}{p_i}\right) \leq p_i \left(\frac{q_i}{p_i} - 1\right) \log_2(e) \Leftrightarrow$$

$$\sum_{i=1}^{n} p_i \log_2\left(\frac{q_i}{p_i}\right) \leq \underbrace{\log_2(e) \sum_{i=1}^{n} (q_i - p_i)}_{=\log_2(e)\,(1-1)=0} \Leftrightarrow \sum_{i=1}^{n} p_i \log_2\left(\frac{q_i}{p_i}\right) \leq 0$$

Σχετική Εντροπία

- Έστω δύο ΣΠΜ p και q, οι οποίες ορίζονται στον ίδιο χώρο πιθανοτήτων (*sample space*) X

- Η σχετική εντροπία (*relative entropy*) ή απόσταση (*divergence*) Kullback-Leibler μεταξύ δύο ΣΠΜ $f(x)$ και $g(x)$ ορίζεται ως

$$\mathcal{D}\left(p||q\right) = \mathbb{E}_p\left\langle \log_2\left[\frac{p(X)}{q(X)}\right]\right\rangle = \sum_{x\in X} p(x)\log_2\left[\frac{p(x)}{q(x)}\right]$$

- Η σχετική εντροπία είναι μέτρο του κατά πόσο μία ΣΠΜ διαφέρει σε σχέση με ΣΠΜ αναφοράς

- Από την ανισότητα Gibbs είναι προφανές ότι

$$\mathcal{D}\left(p||q\right) \geq 0$$

Αμοιβαία Πληροφορία

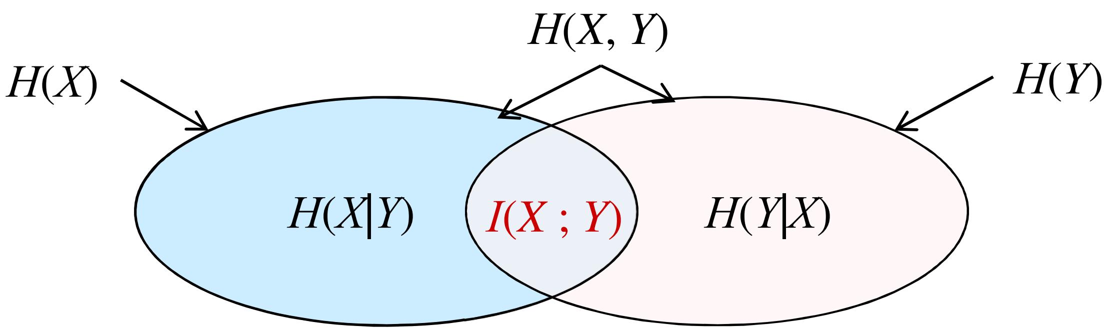
- Βάσει της σχετικής εντροπίας μπορούμε να ορίσουμε την αμοιβαία πληροφορία (*mutual information*) δύο ΤΜ X και Y ως

$$I\,(X;Y) = \mathcal{D}\,(p(X,Y)\;||\;p(X)\,p(Y)) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\,\log_2\left[\frac{p(x_i,y_j)}{p(x_i)\,p(y_j)}\right]$$

- Η $I(X;Y)$ εκφράζει την σχετική εντροπία μεταξύ των κατανομών $p(x_i, y_j)$ και $p(x_i)\,p(y_j)$

- Παρατηρούμε ότι για ανεξάρτητες ΤΜ X και Y, η αμοιβαία πληροφορία μηδενίζεται

- Η $I(X;Y)$ μπορεί να οπτικοποιηθεί με το δίπλα διάγραμμα Venn



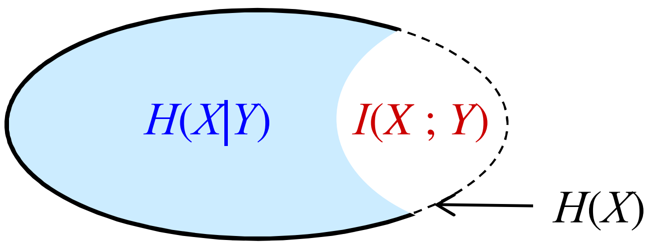
Αμοιβαία Πληροφορία

- Η παραπάνω σχέση μπορεί επίσης να εκφραστεί μέσω της εντροπίας ως

$$I(X;Y) = \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\,\log_2\left[\frac{p(x_i,y_j)}{p(y_j)}\right]}_{-H(X|Y)} \underbrace{-\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\,\log_2\left[p(x_i)\right]}_{H(X)} \Leftrightarrow$$

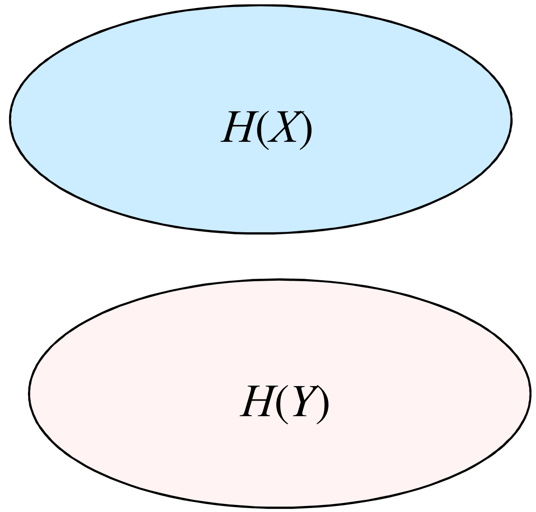
$$I(X;Y) = H(X) - H(X|Y)$$

- Η αμοιβαία πληροφορία εκφράζει την μείωση της αβεβαιότητας της ΤΜ X λόγω γνώσης της Y
  - Αν οι ΤΜ X, Y είναι μεταξύ τους ανεξάρτητες, τότε η γνώση της Y δεν δίνει καμία πληροφορία για την X, δηλαδή $H(X|Y) = H(X)$ και άρα $I(X;Y) = 0$
  - Αν οι ΤΜ X, Y είναι μεταξύ τους πλήρως εξαρτημένες, τότε η γνώση της Y ισοδυναμεί με την απόλυτη γνώση της X, δηλαδή $H(X|Y) = 0$ και άρα $I(X;Y) = H(X)$

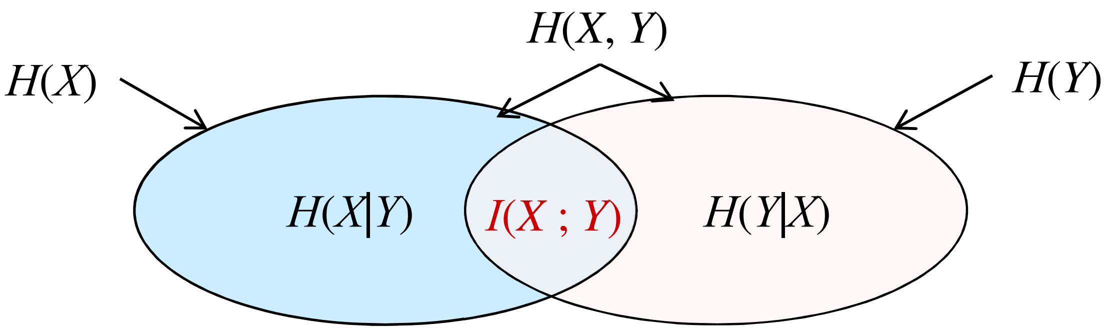


Αμοιβαία Πληροφορία

- Ιδιότητες αμοιβαίας πληροφορίας:
  - $I(X;Y) = H(X) - H(X|Y)$
  - $I(X;Y) = H(Y) - H(Y|X)$
  - $I(X;Y) = H(X) + H(Y) - H(X,Y)$
  - $I(X;Y) = I(Y;X)$
  - $I(X;X) = H(X)$



- Όλες οι παραπάνω ιδιότητες αναπαρίστανται γραφικά στο παρακάτω διάγραμμα Venn



Αμοιβαία Πληροφορία

- Παράδειγμα (*συνέχεια*):
  - Η από κοινού εντροπία βρέθηκε $H(X, Y) = 27/8$ bit
  - Η υπό συνθήκη εντροπία βρέθηκε $H(X|Y) = 11/8$ bit
  - Η εντροπία ως προς την ΤΜ X βρέθηκε $H(X) = 7/4$ bit

		X				
		x_1	x_2	x_3	x_4	$p(y_j)$
Y	y_1	1/8	1/16	1/32	1/32	1/4
	y_2	1/16	1/8	1/32	1/32	1/4
	y_3	1/16	1/16	1/16	1/16	1/4
	y_4	1/4	0	0	0	1/4
	$p(x_i)$	1/2	1/4	1/8	1/8	1

- Η αμοιβαία πληροφορία $I(X; Y)$ μπορεί να υπολογιστεί όπως παρακάτω

$$I(X;Y) = H(X) - H(X|Y) = \left(\frac{7}{4} - \frac{11}{8}\right) \text{ bit} = \frac{3}{8} \text{ bit}$$

Αμοιβαία Πληροφορία

- Βάσει της σχετικής εντροπίας μπορούμε να ορίσουμε και την υπό συνθήκη αμοιβαία πληροφορία (*conditional mutual information*) δύο ΤΜ X και Y, δεδομένης της Z, ως

$$I\left(X;Y|Z\right)=\mathcal{D}\left(p(X,Y|Z)\,||\,p(X|Z)\,p(Y|Z)\right)=$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{p}p(x_i,y_j,z_l)\,\log_2\left[\frac{p(x_i,y_j|z_l)}{p(x_i|z_l)\,p(y_j|z_l)}\right]$$

- Η $I(X;Y|Z)$ εκφράζει την σχετική εντροπία μεταξύ των υπό συνθήκη κατανομών $p(x_i, y_j | z_l)$ και $p(x_i | z_l)\, p(y_j | z_l)$

Αμοιβαία Πληροφορία

- Η υπό συνθήκη αμοιβαία πληροφορία μπορεί να γραφεί ως

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

- <u>Απόδειξη</u>: Από τον ορισμό της υπό συνθήκη αμοιβαίας πληροφορίας έχουμε

$$I\left(X;Y|Z\right) = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{\ell} p(x_i, y_j, z_l)\ \log_2\left[\frac{p(x_i, y_j|z_l)}{\boxed{p(x_i|z_l)}\,p(y_j|z_l)}\right] =$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{\ell} p(x_i, y_j, z_l)\ \log_2\left[p(x_i|z_l)\right] + \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{\ell} p(x_i, y_j, z_l)\ \log_2\left[\frac{p(x_i, y_j|z_l)}{p(y_j|z_l)}\right] =$$

$$= \underbrace{-\sum_{i=1}^{n}\sum_{l=1}^{\ell} p(x_i, z_l)\ \log_2\left[p(x_i|z_l)\right]}_{H(X|Z)} + \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{\ell} p(x_i, y_j, z_l)\ \log_2\left[p(x_i|y_j, z_l)\right]}_{-H(X|Y,Z)}$$

Αμοιβαία Πληροφορία

- Κανόνας αλυσίδας της αμοιβαίας πληροφορίας: Η αμοιβαία πληροφορία γράφεται

$$I\left(X_1, X_2, \ldots, X_n; Y\right) = \sum_{i=1}^{n} I\left(X_i; Y | X_{i-1}, X_{i-2}, \ldots, X_1\right)$$

- Η $I(X_1,X_2,\ldots,X_n;Y)$ είναι η τομή μεταξύ της από κοινού εντροπίας $H(X_1, X_2,\ldots,X_n)$ και της $H(Y)$

- Απόδειξη: Η αμοιβαία πληροφορία εκφράζει την μείωση της αβεβαιότητας της των από κοινού ΤΜ X_1, $X_2,\ldots,X_n$ λόγω της γνώσης της Y

$$I\left(X_1, X_2, \ldots, X_n; Y\right) = H(X_1, X_2, \ldots, X_n) - H(X_1, X_2, \ldots, X_n | Y) =$$

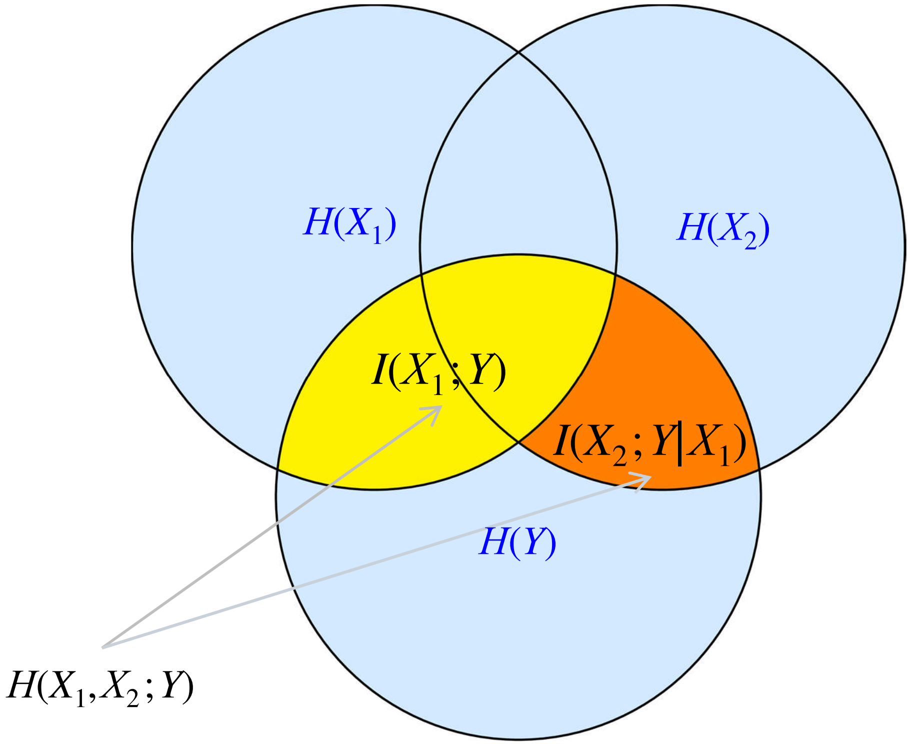
$$= \sum_{i=1}^{n} H(X_i | X_{i-1}, X_{i-2}, \ldots, X_1) - \sum_{i=1}^{n} H(X_i | X_{i-1}, X_{i-2}, \ldots, X_1, Y) =$$

$$= \sum_{i=1}^{n} \underbrace{\left[H(X_i | X_{i-1}, X_{i-2}, \ldots, X_1) - H(X_i | X_{i-1}, X_{i-2}, \ldots, X_1, Y)\right]}_{I(X_i;Y|X_{i-1},X_{i-2},\ldots,X_1)} =$$

Αμοιβαία Πληροφορία

- Παράδειγμα εφαρμογής του κανόνα της αλυσίδας για *n* = 2 ΤΜ

$$I\left(X_1, X_2; Y\right) = \sum_{i=1}^{2} I\left(X_i; Y | X_{i-1}\right) = I(X_1; Y) + I\left(X_2; Y | X_1\right)$$



Αμοιβαία Πληροφορία

- Παράδειγμα εφαρμογής του κανόνα της αλυσίδας για n = 3 ΤΜ

$$I\left(X_1, X_2, X_3; Y\right) = \sum_{i=1}^{3} I\left(X_i; Y \middle| X_{i-1}, X_{i-2}\right) =$$
$$= I(X_1; Y) + I\left(X_2; Y \middle| X_1\right) + I\left(X_3; Y \middle| X_2, X_1\right)$$