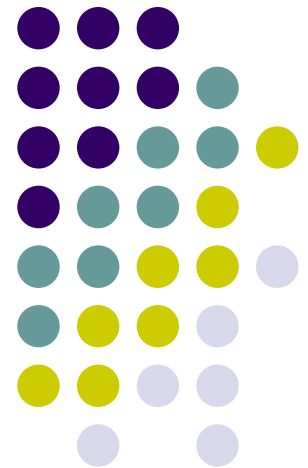
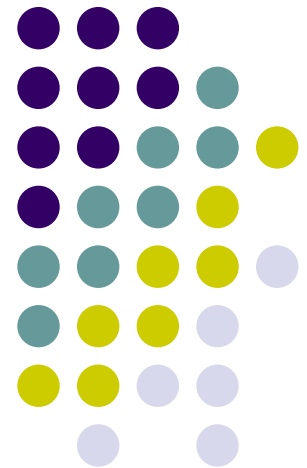


Θεωρία πληροφορίας



Εισαγωγή στη θεωρία πληροφορίας

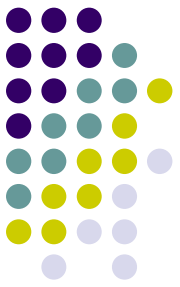


Τηλεπικοινωνιακά συστήματα



- Όλα τα τηλεπικοινωνιακά συστήματα σχεδιάζονται για να μεταφέρουν πληροφορία
 - Σε κάθε τηλεπικοινωνιακό σύστημα υπάρχει μια πηγή που παράγει πληροφορία
 - Σκοπός του τηλεπικοινωνιακού συστήματος είναι να μεταφέρει την πληροφορία στον παραλήπτη
- Π.χ. στην τηλεοπτική εκπομπή
 - Η πηγή πληροφορίας είναι η βιντεοκάμερα
 - Η έξοδος της είναι η εικόνα
- Σε επικοινωνία υπολογιστών
 - Η πηγή πληροφορίας είναι κάποια εφαρμογή π.χ. email
 - Η έξοδος είναι χαρακτήρες ASCII

Πληροφορία – κοινή αντίληψη



- Κατά την κοινή αντίληψη πληροφορία = νέα γνώση
 - Αποκτάται μέσω ακοής, όρασης, κλπ
- Η πηγή πληροφορίας παράγει εξόδους που δεν είναι γνωστές στον δέκτη εκ των προτέρων
 - Εάν μπορούσαμε να τις προβλέψουμε δεν θα υπήρχε λόγος μετάδοσης
- Τι είναι πληροφορία;
 - Η ποιοτική περιγραφή (γνώση για κάποιο θέμα) δεν είναι αρκετή
 - Απαιτείται ποσοτική περιγραφή

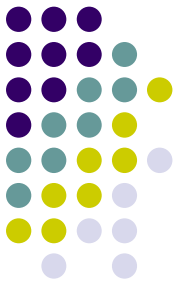
Πληροφορία – διαισθητική προσέγγιση



- Υποθέστε ότι στις σημερινές εφημερίδες έχουμε τα ακόλουθα πρωτοσέλιδα
 - Αύριο ο ήλιος θα ανατείλει
 - Καίγεται η Ελλάδα
 - Καίγεται η Ελβετία
 - Εισβολή της Κούβας στις ΗΠΑ
- Πληροφορία \rightarrow Έκπληξη \rightarrow 1/Πιθανότητα

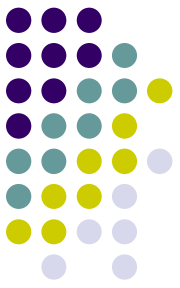
Πρωτοσέλιδο	Έκπληξη	Πιθανότητα	Πληροφορία
1	Καμία	Πάντα	Όχι
2	Μικρή	Μεγάλη	Αρκετή
3	Μεγάλη	Μικρή	Μεγάλη
4	Τεράστια	Ελάχιστη	Τεράστια

Πληροφορία – προσέγγιση μηχανικού



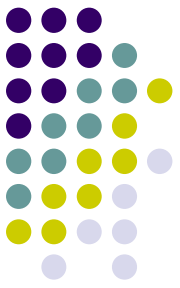
- Η πληροφορία σε ένα μήνυμα είναι ο χρόνος που απαιτείται για τη μετάδοσή του
 - Μηνύματα με μεγαλύτερη πιθανότητα εμφάνισης μπορούν να μεταδοθούν σε συντομότερο χρόνο από ότι πιο σπάνια μηνύματα
- Κώδικας Morse
 - Μικρότερες κωδικές λέξεις για χαρακτήρες που εμφανίζονται πιο συχνά (α, ε, ο)
 - Μεγαλύτερες κωδικές λέξεις για χαρακτήρες που εμφανίζονται πιο σπάνια (ξ, ψ, ζ)
- Ο χρόνος μετάδοσης ενός χαρακτήρα με πιθανότητα εμφάνισης p είναι ανάλογος του $\log(1/p)$

Μοντέλα πηγών πληροφορίας

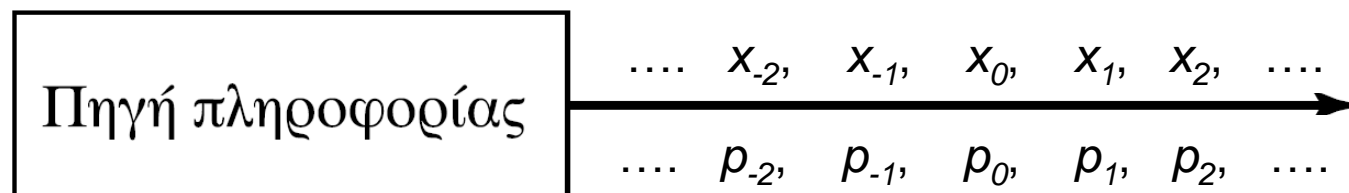


- Οι πηγές πληροφορίας μπορούν να μοντελοποιηθούν με τυχαίες διαδικασίες
 - Η έξοδος μιας πηγής είναι χρονικά μεταβαλλόμενη συνάρτηση (χωρίς δυνατότητα πρόβλεψης)
 - Οι ιδιότητες εξαρτώνται από τη φύση της πηγής
- Κοινός τόπος είναι ότι όλες οι πηγές έχουν πεπερασμένο εύρος ζώνης, άρα μπορούν να δειγματοληπτηθούν με τον ρυθμό Nyquist
 - Περιοριζόμαστε σε διακριτές ως προς το χρόνο διαδικασίες
$$\{X_i\}_{i=-\infty}^{\infty}$$
- Το αλφάβητο των X_i μπορεί να είναι διακριτό (π.χ. χαρακτήρες) ή συνεχές (π.χ. δείγματα φωνής)

Διακριτή πηγή χωρίς μνήμη (Discrete Memoryless Source)



- Το πιο απλό μοντέλο πηγής
- Η DMS είναι στοχαστική ανέλιξη διακριτού χρόνου και διακριτού πλάτους
 - Τα x_i δημιουργούνται ανεξάρτητα, το καθένα με τη δική του πιθανότητα p_i
 - Εάν χρησιμοποιείται κβάντιση, τα συνεχή σήματα μπορούν να θεωρηθούν και αυτά ως διακριτά
 - Πιο πολύπλοκα μοντέλα έχουν μνήμη, δηλ. η δημιουργία κάποιου συμβόλου εξαρτάται από τα προηγούμενα





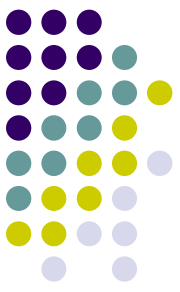
Μέτρο πληροφορίας

- Ορίζεται με τέτοιο τρόπο ώστε να ικανοποιούνται οι διαισθητικές ιδιότητες
- Για κάθε έξοδο της πηγής υπάρχει μια συνδεδεμένη πιθανότητα εμφάνισής της
 - Ποια έξοδος μεταφέρει περισσότερη πληροφορία;
- Το μέτρο πληροφορίας πρέπει να είναι φθίνουσα συνάρτηση της πιθανότητας
 - Όσο πιο απίθανο το γεγονός, τόσο πιο μεγάλη πληροφορία μεταφέρει



Μέτρο πληροφορίας

- Το μέτρο πληροφορίας πρέπει να είναι συνεχής συνάρτηση
 - Μικρές αλλαγές της πιθανότητας εμφάνισης της εξόδου δεν πρέπει να μεταβάλλουν πολύ την ποσότητα της πληροφορίας
- Αθροιστική ιδιότητα
 - Η συνολική ποσότητα πληροφορίας δύο ανεξάρτητων γεγονότων είναι το άθροισμα των επί μέρους μέτρων πληροφορίας των γεγονότων



Ιδία-πληροφορία

- Το περιεχόμενο πληροφορίας εξαρτάται μόνο από την πιθανότητα p_j εμφάνισης και όχι από το μέγεθος του συμβόλου x_j
 - *Self-information* $I(p_j)$
- Η ίδια-πληροφορία $I(p_j)$ είναι συνεχής συνάρτηση της p_j
- Η ίδια-πληροφορία $I(\cdot)$ είναι φθίνουσα συνάρτηση του ορίσματος
- Εάν $p = p_1 p_2$, τότε $I(p) = I(p_1) + I(p_2)$



Μέτρο πληροφορίας

- Η μόνη συνάρτηση που ικανοποιεί τις προηγούμενες ιδιότητες είναι
 - $I(x) = -\log(p(x))$
- Η βάση του λογαρίθμου δεν είναι σημαντική, αλλά ορίζει τη μονάδα μέτρησης της πληροφορίας
- Εάν η βάση είναι **2**, τότε η πληροφορία μετριέται σε **bit**
- Εάν χρησιμοποιηθούν φυσικοί λογάριθμοι, τότε η πληροφορία μετριέται σε **nap**
- Στη συνέχεια όλοι οι λογάριθμοι θα είναι με βάση το **2**



Σημαντικές ιδιότητες

- Προφανώς ισχύει

$$I(x_i) = 0 \quad \text{εάν} \quad p(x_i) = 1$$

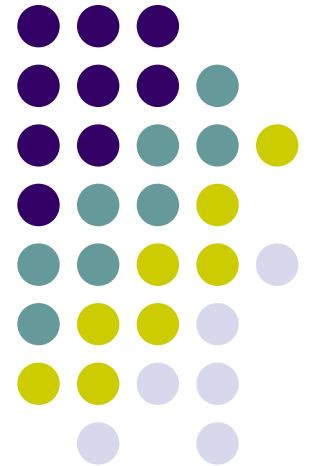
$$I(x_i) \geq 0$$

$$I(x_i) > I(x_j) \quad \text{εάν} \quad p(x_i) < p(x_j)$$

$$I(x_i x_j) = I(x_i) + I(x_j) \quad \text{εάν} \quad x_i, x_j \text{ ανεξάρτητα}$$

- Τι συμβαίνει όταν $p_i=0$;

Εντροπία

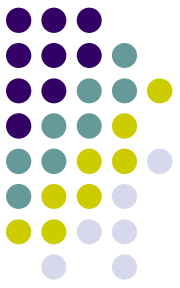




Μέση πληροφορία

- Τηλεπικοινωνιακό σύστημα
 - Μεταδίδονται μακριές σειρές συμβόλων
- Η ποσότητα πληροφορίας πηγής είναι η μέση τιμή της ίδιας-πληροφορίας όλων των εξόδων της

$$\sum_{i=1}^N p_i I(x_i) = - \sum_{i=1}^N p_i \log(p_i)$$



Εντροπία

- **Εντροπία $H(X)$** είναι η μέση τιμή πληροφορίας της πηγής ανά **σύμβολο**

$$H(X) = E \{ [I(X)] \} = - \sum_{i=1}^N p_i \log(p_i) = \sum_{i=1}^N p_i \log\left(\frac{1}{p_i}\right)$$

- Η εντροπία αποτελεί ένα μέτρο της **αβεβαιότητας** για την x (κατά μέσο όρο)
 - Όσο πιο πολλά είναι γνωστά για την x , τόσο μικρότερη είναι η εντροπία

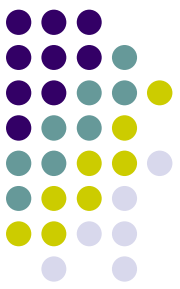
Εντροπία ομοιόμορφης κατανομής



- Εάν οι έξοδοι της πηγής είναι ισοπίθανες, τότε

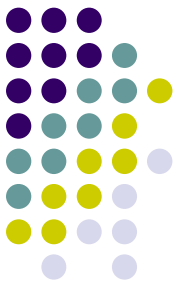
$$x_i, i = 1, 2, \dots, N \rightarrow p_i = 1/n$$

$$H(X) = \sum_{i=1}^N \frac{1}{N} \log(N) = \log(N)$$



Απλά όρια

- Για μια τυχαία μεταβλητή με N πιθανές τιμές
 $0 \leq H(X) \leq \log(N)$
- Πότε επιτυγχάνονται τα όρια;
 - $H(X) = 0$ εάν και μόνο εάν $p(x_i) = 1$ για κάποιο i
 - $H(X) = \log(N)$ εάν και μόνο εάν $p(x_i) = 1/N$ για όλα τα i



Δυαδική πηγή χωρίς μνήμη

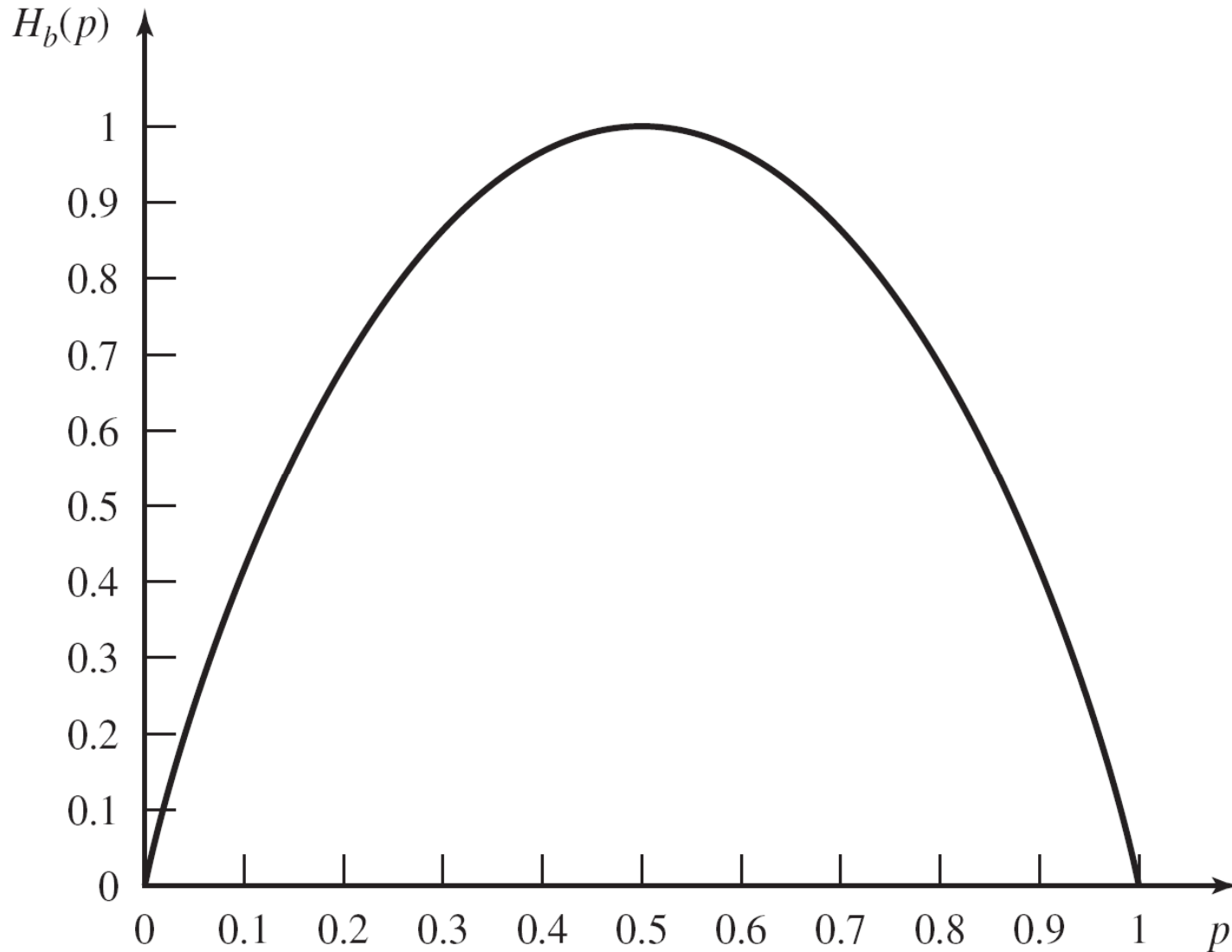
- Εμφανίζονται δύο πιθανές έξοδοι με πιθανότητες p και $1-p$, αντίστοιχα

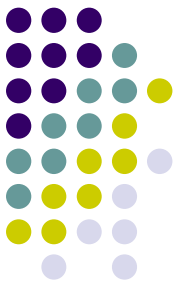
- Η εντροπία είναι

$$\begin{aligned} H(X) &= -p \log(p) - (1-p) \log(1-p) \\ &= H_b(p) \end{aligned}$$

- Μεγιστοποιείται όταν $p=1/2$, δηλαδή, $H_b(1/2)=1$
 - Όπως αναμένεται το αποτέλεσμα μπορεί να μεταφερθεί με 1 bit

Συνάρτηση δυαδικής εντροπίας





Από κοινού εντροπία

- Η από κοινού εντροπία δύο τυχαίων μεταβλητών X, Y είναι

$$H(X, Y) = - \sum_{x, y} p(x, y) \log p(x, y)$$

- Στη γενική περίπτωση n τυχαίων μεταβλητών

$$H(\mathbf{X}) = - \sum_{x_1, x_2, \dots, x_n} p(x_1, x_2, \dots, x_n) \log p(x_1, x_2, \dots, x_n)$$



Υπό συνθήκη εντροπία

- Εάν η τιμή της τυχαίας μεταβλητής $Y=y$ είναι γνωστή, η συνάρτηση μάζας πιθανότητας της X θα είναι $p(x|y)$ και η αντίστοιχη εντροπία είναι

$$H(X|Y=y) = -\sum_{x,y} p(x|y) \log p(x|y)$$

- Η ποσότητα αυτή αντιστοιχεί στην **αβεβαιότητα** της X όταν γνωρίζουμε ότι $Y=y$



Υπό συνθήκη εντροπία

- Η υπό συνθήκη εντροπία είναι μέτρο για την **αβεβαιότητα** της X δοθείσης της Y

$$H(X | Y) = - \sum_{x,y} p(x | y) p(y) \log p(x | y)$$

$$= - \sum_{x,y} p(x, y) \log p(x | y)$$

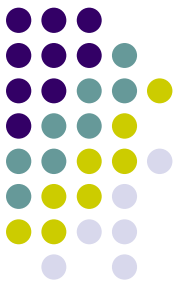
- δηλαδή, είναι η μέση τιμή για όλα τα y για της αβεβαιότητας για τη X όταν η Y είναι γνωστή
- Επειδή η Y μπορεί να παρέχει κάποια πληροφορία για την X , ισχύει $H(X | Y) \leq H(X)$



Υπό συνθήκη εντροπία

- Στη γενική περίπτωση έχουμε

$$H(X_n | X_1, X_2, \dots, X_{n-1}) =$$
$$- \sum_{x_1, x_2, \dots, x_n} p(x_1, x_2, \dots, x_n) \log p(x_n | x_1, x_2, \dots, x_{n-1})$$



Ρυθμός εντροπίας

- Εάν X_n είναι η έξοδος μιας διακριτής (όχι απαραίτητα χωρίς μνήμη) πηγής

$$H(X_n | X_1, X_2, \dots, X_{n-1})$$

είναι η καινούργια πληροφορία για τη X_n όταν κάποιος έχει παρακολουθήσει την ακολουθία X_1, X_2, \dots, X_{n-1}

- Το όριο H στο άπειρο είναι ο **ρυθμός εντροπίας**



Ρυθμός εντροπίας

- Η στατικότητα εξασφαλίζει την ύπαρξη του ορίου

$$H = \lim_{n \rightarrow \infty} H(X_n | X_1, X_2, \dots, X_{n-1})$$

και ισχύει εναλλακτικά

$$H = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n)$$

ακόμη και για πηγές με μνήμη



Χρήσιμες ιδιότητες

- Αλυσίδα

$$\begin{aligned} H(X_1, X_2, \dots, X_n) &= H(X_1) + \\ &\quad H(X_2 | X_1) + \dots + \\ &\quad H(X_n | X_1, X_2, \dots, X_{n-1}) \end{aligned}$$

- Για δύο τυχαίες μεταβλητές

$$\begin{aligned} H(X, Y) &= H(Y) + H(X | Y) \\ &= H(X) + H(Y | X) \end{aligned}$$



Χρήσιμες ιδιότητες

- Εάν οι τυχαίες μεταβλητές είναι ανεξάρτητες

$$H(\mathbf{X}) = \sum_{i=1}^n H(X_i)$$

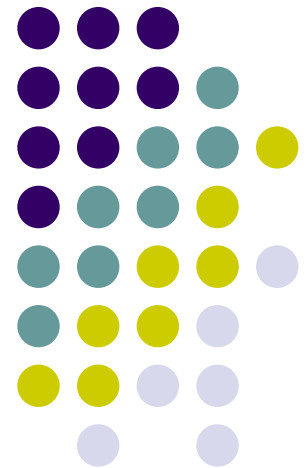
και εάν έχουν την ίδια κατανομή

$$H(\mathbf{X}) = nH(X_1)$$

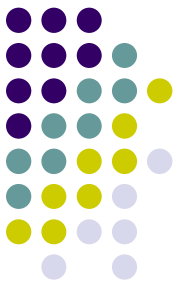
- Ισχύει η ανισότητα (ισότητα όταν οι τυχαίες μεταβλητές είναι ανεξάρτητες)

$$H(\mathbf{X}) \leq \sum_{i=1}^n H(X_i)$$

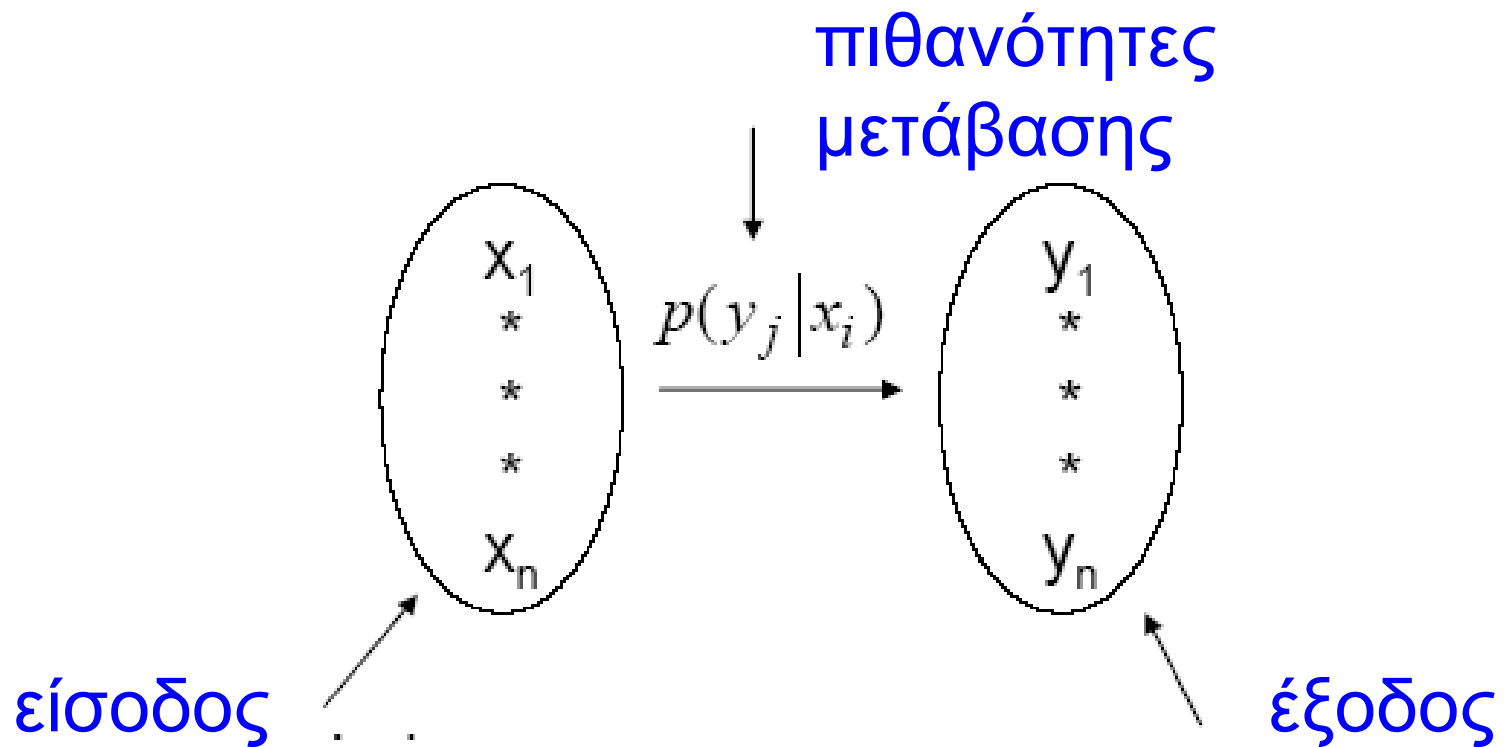
Αμοιβαία πληροφορία



Μοντέλο τηλεπικοινωνιακού διαύλου



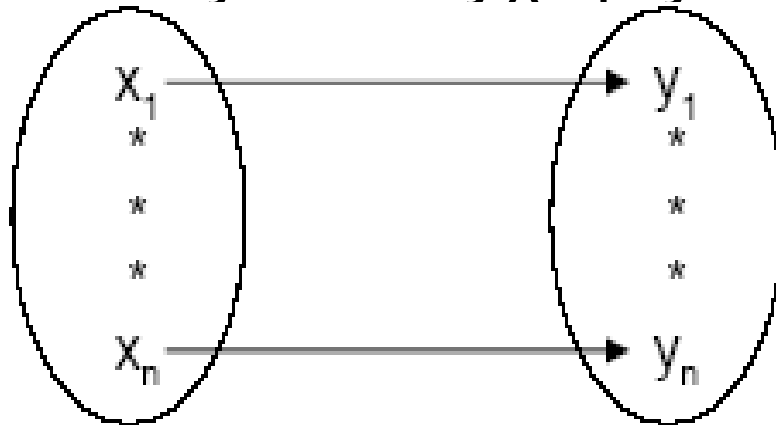
- Μοντέλο διακριτού διαύλου



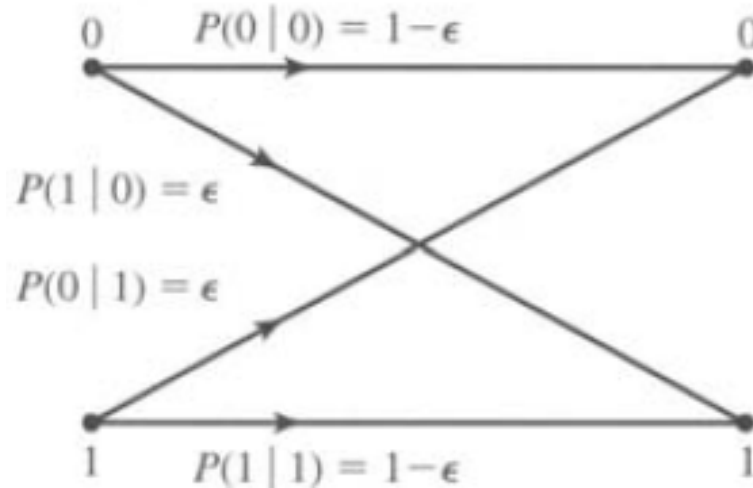


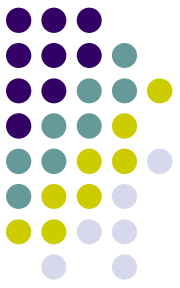
Παραδείγματα διαύλων

- Ιδανικός δίαυλος χωρίς θόρυβο



- Δυαδικός συμμετρικός δίαυλος

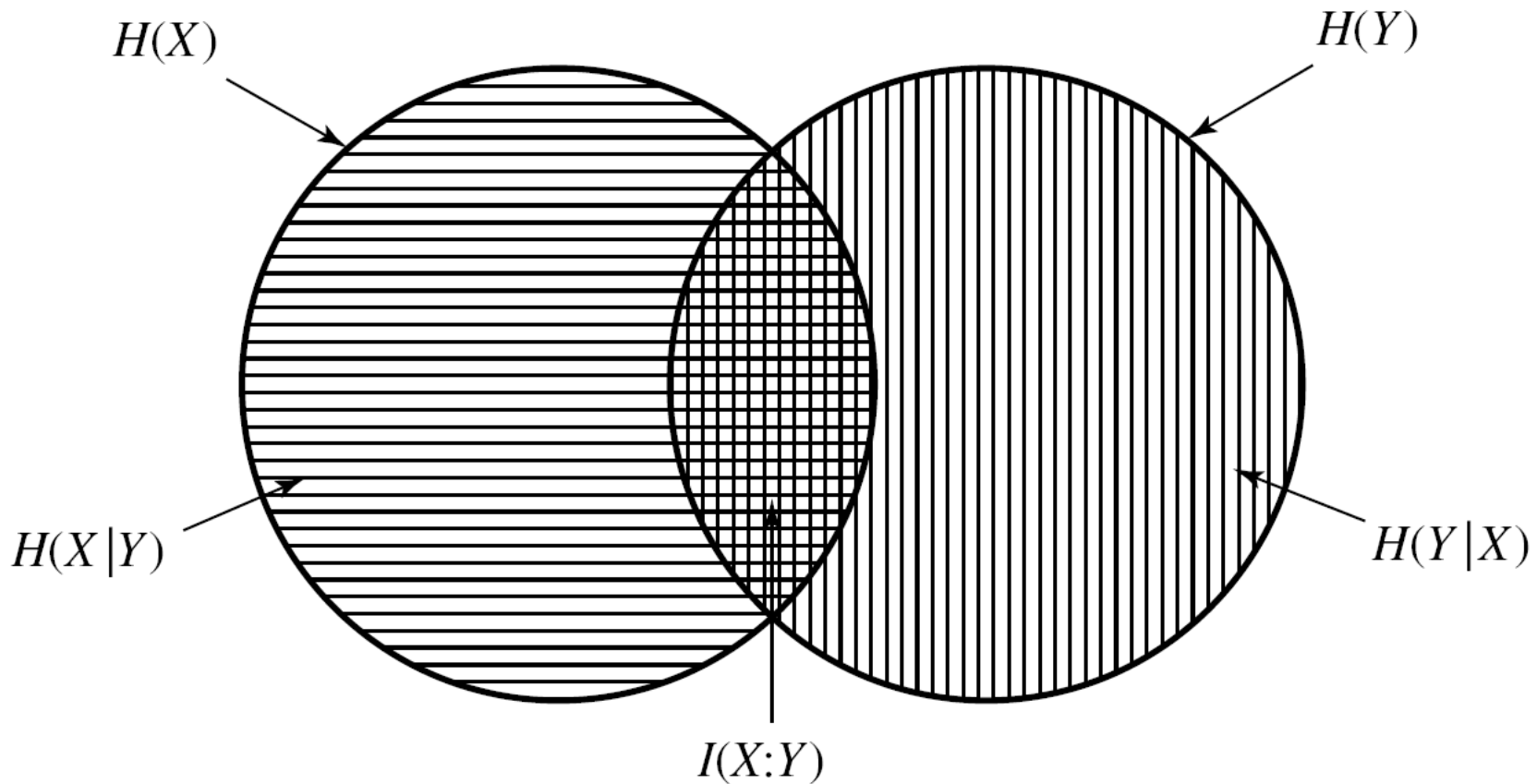




Αμοιβαία πληροφορία

- Η αμοιβαία πληροφορία $I(X; Y)$ μεταξύ δύο τυχαίων μεταβλητών είναι
$$I(X; Y) = H(X) - H(X | Y)$$
 - Μπορούμε να θεωρήσουμε την X ως την είσοδο και την Y ως την έξοδο του διαύλου
- $H(X)$ είναι η αβεβαιότητα για την X
- $H(X | Y)$ είναι η αβεβαιότητα για την X δοθέντος ότι γνωρίζουμε την Y
- $H(X) - H(X | Y)$ είναι η αβεβαιότητα για την X που εξαλείφεται με την εμφάνιση της Y

Εντροπία, υπό συνθήκη εντροπία και αμοιβαία πληροφορία



Ιδιότητες



- Ισχύει ότι
$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) - (H(X,Y) - H(Y)) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned}$$
- άρα $I(X;Y) = I(Y;X)$
- $I(X;Y) \geq 0$ επειδή $H(Y) \geq H(Y|X)$
 - με την ισότητα να ισχύει εάν οι τυχαίες μεταβλητές είναι ανεξάρτητες
- $I(X;Y) \leq \min\{H(X), H(Y)\}$



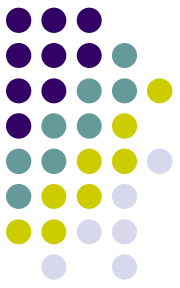
Διαφορική εντροπία

- Για τυχαία μεταβλητή X με συνεχείς τιμές, η εντροπία είναι άπειρη
- Στην περίπτωση αυτή ορίζεται η διαφορική εντροπία

$$h(X) = - \int_{-\infty}^{\infty} f_X(x) \log f_X(x) dx$$

- Η διαφορική εντροπία δεν έχει τη δαισθητική ερμηνεία της διακριτής εντροπίας, όμως είναι χρήσιμη για τη σύγκριση της πληροφορίας συνεχών πηγών

Απόλυτη εντροπία



- Η απόλυτη εντροπία είναι

$$H_{abs}(X) = - \lim_{\Delta x \rightarrow 0} \sum_i f_X(x_i) \Delta x \log(f_X(x_i) \Delta x)$$

$$= - \lim_{\Delta x \rightarrow 0} \left[\sum_i f_X(x_i) \log f_X(x_i) + f_X(x_i) \log \Delta x \right] \Delta x$$

$$= - \int_{-\infty}^{\infty} f_X(x) \log f_X(x) dx + H_0(X) = h(X) + H_0(X)$$

- **ΌΤΠΟΥ**

$$H_0(X) = -\log 0 = \infty$$

Αμοιβαία πληροφορία συνεχών τυχαίων μεταβλητών



- Για δύο τυχαίες μεταβλητές

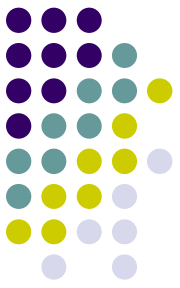
$$h(X, Y) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \log f(x, y) dx dy$$

$$h(X | Y) = h(X, Y) - h(Y)$$

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

- όπου η αμοιβαία πληροφορία των συνεχών τυχαίων μεταβλητών έχει τη γνωστή ερμηνεία της πληροφορίας που η μία παρέχει για την άλλη

Διαφορική εντροπία Γκαουσιανής τυχαίας μεταβλητής



- Η συνάρτηση πυκνότητας πιθανότητας είναι

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

- άρα

$$h(X) = -\int_{-\infty}^{\infty} f(x) \log \frac{1}{\sqrt{2\pi\sigma^2}} dx - \int_{-\infty}^{\infty} f(x) \log \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

$$= -\log \frac{1}{\sqrt{2\pi\sigma^2}} + \frac{1}{2\sigma^2} \log e \int_{-\infty}^{\infty} x^2 f(x) dx$$

$$= \frac{1}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sigma^2 \log e = \frac{1}{2} \log(2\pi e \sigma^2)$$