EE728
Προχωρημένα Θέματα Θεωρίας Πληροφορίας
2η διάλεξη
(4η έκδοση, 17/3/2011)

Δημήτρης-Αλέξανδρος Τουμπακάρης

Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών

1 Μαρτίου 2011

Περιεχόμενα 2ης διάλεξης

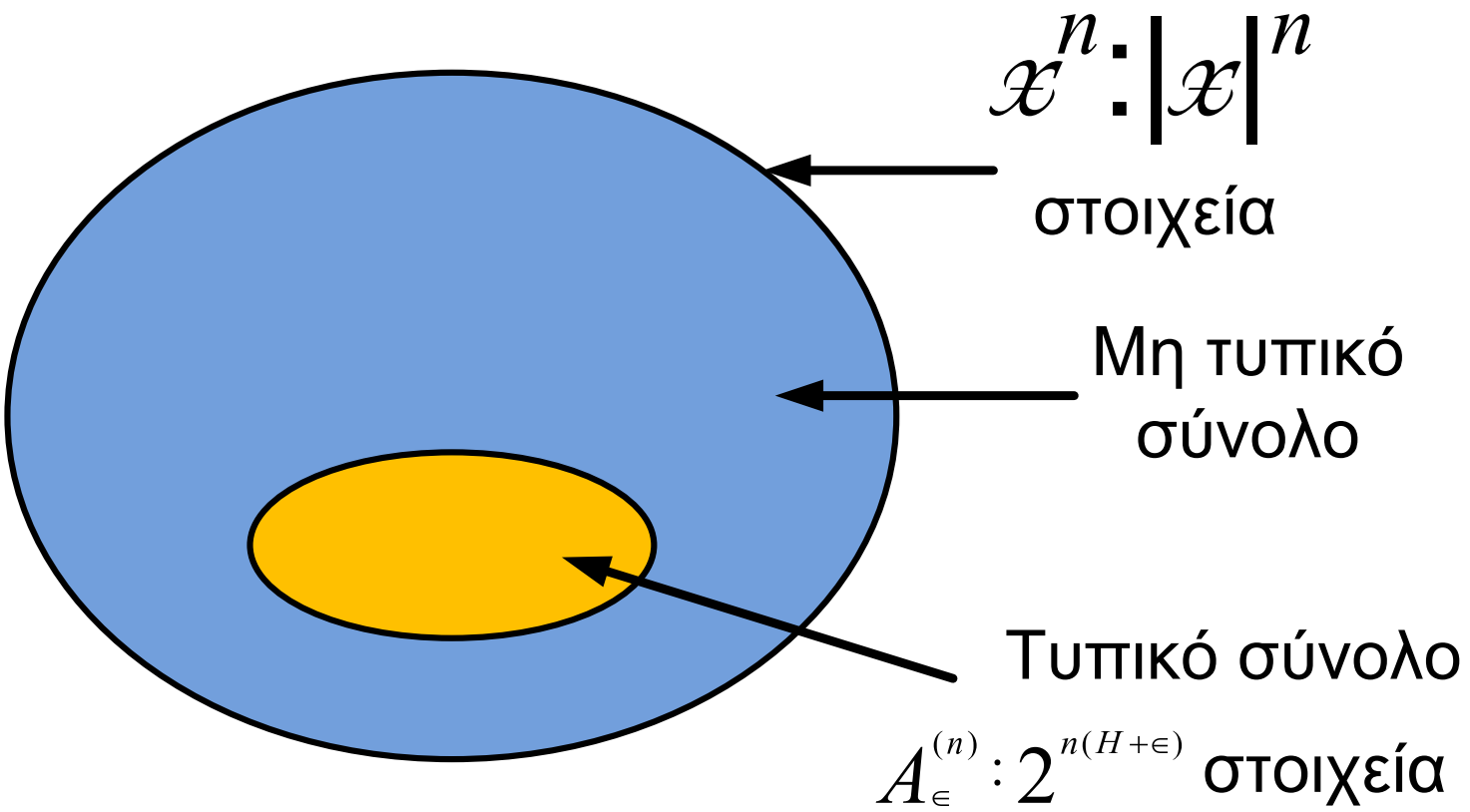
Τυπικό Σύνολο (Typical Set) και ιδιότητες

- Ορισμός 2.1 Το (ασθενώς) τυπικό σύνολο (weakly typical set) $A_\epsilon^{(n)}$ που αντιστοιχεί στην κατανομή $p(x)$ αποτελείται από τις ακολουθίες $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ που ικανοποιούν τη σχέση

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, x_2, \ldots, x_n) \leq 2^{-n(H(X)-\epsilon)}.$$

- Ιδιότητες $A_\epsilon^{(n)}$:
  1. Εάν $(x_1, x_2, \ldots, x_n) \in A_\epsilon^{(n)}$,
     $H(X) - \epsilon \leq -\frac{1}{n} \log p(x_1, x_2, \ldots, x_n) \leq H(X) + \epsilon.$
  2. $\Pr\left\{A_\epsilon^{(n)}\right\} > 1 - \epsilon$ για n μεγαλύτερο από κάποια τιμή n_0.
  3. $\left|A_\epsilon^{(n)}\right| \leq 2^{n(H(X)+\epsilon)}$,
     όπου $\left|A\epsilon^{(n)}\right|$ ο αριθμός των στοιχείων του τυπικού συνόλου $A_\epsilon^{(n)}$.
  4. $\left|A_\epsilon^{(n)}\right| \geq (1-\epsilon)2^{n(H(X)-\epsilon)}$, για n μεγαλύτερο από κάποια τιμή n_0.

Τυπικό Σύνολο



Αποδείξεις ιδιοτήτων Τυπικού Συνόλου

1. Εάν $(x_1, x_2, \ldots, x_n) \in A_\epsilon^{(n)}$,
   $H(X) - \epsilon \leq -\frac{1}{n} \log p(x_1, x_2, \ldots, x_n) \leq H(X) + \epsilon$.
   Προκύπτει άμεσα από τον ορισμό του τυπικού συνόλου παίρνοντας το λογάριθμο.
2. $\Pr\left\{A_\epsilon^{(n)}\right\} > 1 - \epsilon$ για n μεγαλύτερο από κάποια τιμή n_0.
   Προκύπτει άμεσα από το AEP δεδομένου ότι η πιθανότητα μια ακολουθία να είναι τυπική τείνει στο 1 καθώς το n τείνει στο άπειρο. Επομένως, για κάθε $\delta > 0$, υπάρχει n_0 τέτοιο ώστε, για $n \geq n_0$,

$$\Pr\left\{\left|-\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) - H(X)\right| < \epsilon\right\} > 1 - \delta.$$

   Θέτοντας $\delta = \epsilon$ προκύπτει η ιδιότητα.

Αποδείξεις ιδιοτήτων Τυπικού Συνόλου (2)

3. $\left|A_\epsilon^{(n)}\right| \leq 2^{n(H(X)+\epsilon)}$.

$$1 = \sum_{\mathbf{x}\in\mathcal{X}^n} p(\mathbf{x}) \geq \sum_{\mathbf{x}\in A_\epsilon^{(n)}} p(\mathbf{x}) \overset{(a)}{\geq} \sum_{\mathbf{x}\in A_\epsilon^{(n)}} 2^{-n(H(X)+\epsilon)}$$
$$= 2^{-n(H(X)+\epsilon)} \left|A_\epsilon^{(n)}\right|.$$

Στο (a) χρησιμοποιήθηκε ο ορισμός του τυπικού συνόλου.

4. $\left|A_\epsilon^{(n)}\right| \geq (1-\epsilon)2^{n(H(X)-\epsilon)}$, για n μεγαλύτερο από κάποια τιμή n_0.

Από τη 2η ιδιότητα, για $n \geq n_0$,

$$1-\epsilon < \Pr\left\{A_\epsilon^{(n)}\right\} = \sum_{\mathbf{x}\in A_\epsilon^{(n)}} p(\mathbf{x}) \leq \sum_{\mathbf{x}\in A_\epsilon^{(n)}} 2^{-n(H(X)-\epsilon)}$$
$$= 2^{-n(H(X)-\epsilon)} \left|A_\epsilon^{(n)}\right|.$$

Αποδείξεις ιδιοτήτων Τυπικού Συνόλου (3)

- Μπορεί, επίσης, να αποδειχτεί ότι υπάρχει ϵ' τέτοιο ώστε, για n μεγαλύτερο απο κάποια τιμή n_0',

$$\left|A_\epsilon^{(n)}\right| \geq 2^{n(H(X)-\epsilon')}.$$

- Δείτε π.χ. El Gamal & Kim, *Lecture Notes on Information Theory*, Ch. 2.

Παράδειγμα 2.1 (Cover & Thomas Problem 3.6)

- Έστω οι ανεξάρτητες και ομοίως κατανεμημένες τ.μ. $X_1, X_2, \ldots, X_n$ που ακολουθούν κατανομή $p(x)$. Να βρεθεί η τιμή του ορίου

$$\lim_{n\to\infty} \left\{ p(X_1, X_2, \ldots, X_n)^{1/n} \right\}.$$

- Απάντηση:

$$\lim_{n\to\infty} \left\{ \log p(X_1, X_2, \ldots, X_n)^{1/n} \right\} = \lim_{n\to\infty} \left\{ \frac{1}{n} \log p(X_1, X_2, \ldots, X_n) \right\}$$

$$\Rightarrow \lim_{n\to\infty} \left\{ p(X_1, X_2, \ldots, X_n)^{1/n} \right\} = 2^{-H(X)}.$$

Σχέση τυπικού συνόλου με σύνολα που περιέχουν σχεδόν όλη την πιθανότητα

- Είδαμε ότι (ιδιότητα 2), $\Pr\left\{A_\epsilon^{(n)}\right\} > 1-\epsilon$ για n μεγαλύτερο από κάποια τιμή n_0.
- Ένα ερώτημα που δεν έχει απαντηθεί ακόμη είναι το εξής: Μήπως υπάρχει κάποιο σύνολο τέτοιο ώστε $\Pr\left\{B_\epsilon^{(n)}\right\} > 1-\epsilon$ και $\left|B_\epsilon^{(n)}\right| < \left|A_\epsilon^{(n)}\right|$;
- Μήπως, δηλαδή, μπορούμε να ελαττώσουμε περαιτέρω τον αριθμό ακολουθιών που κωδικοποιούμε;
- Αποδεικνύεται (δείτε π.χ. Cover & Thomas Theorem 3.3.1) ότι το τυπικό σύνολο, $A_\epsilon^{(n)}$, έχει περίπου το ίδιο μέγεθος με το μικρότερο σύνολο, $B_\epsilon^{(n)}$, που περιέχει σχεδόν όλη την πιθανότητα.

Ισχυρή Τυπικότητα (Strong Typicality)

- Έως τώρα ασχοληθήκαμε με την ασθενή τυπικότητα.
- Μια ακολουθία είναι ασθενώς τυπική όταν η εμπειρική της εντροπία βρίσκεται κοντά στην πραγματική εντροπία της πηγής που παράγει την ακολουθία.
- Για να είναι μια ακολουθία ισχυρώς τυπική πρέπει η σχετική συχνότητα με την οποία εμφανίζεται κάθε σύμβολο μέσα στην ακολουθία να βρίσκεται κοντά στην κατανομή της πηγής.
- Για παράδειγμα, για πηγή Bern$(1/2)$, η ακολουθία $0\ 0\ 0\ 1\ 0\ 0\ 0$ είναι ασθενώς τυπική, αλλά όχι ισχυρώς τυπική. Η ακολουθία $0\ 0\ 0\ 1\ 1\ 0\ 1\ 1$ είναι ισχυρώς και ασθενώς τυπική.

Ισχυρώς Τυπικό Σύνολο – ορισμός

- Θεωρούμε πηγή χωρίς μνήμη με κατανομή $p(x)$. Έστω ότι $\mathcal{S}_X \subseteq \mathcal{X}$ είναι το σύνολο στο οποίο $p(x) > 0$.
- Το ισχυρώς τυπικό σύνολο $\mathcal{T}_\epsilon^{(n)}$ που αντιστοιχεί στην κατανομή $p(x)$ αποτελείται από τις ακολουθίες $X_1^n \in \mathcal{X}^n$ για τις οποίες $N(x; X_1^n) = 0$ για $x \notin \mathcal{S}_X$ και

$$\sum_{x \in \mathcal{S}_X} \left| \frac{1}{n} N(x; X_1^n) - p(x) \right| \leq \epsilon,$$

όπου $N(x; X_1^n)$ είναι ο αριθμός των εμφανίσεων του στοιχείου x μέσα στην ακολουθία X_1^n και ϵ είναι αυθαίρετα μικρός πραγματικός αριθμός.

- Οι ακολουθίες που ανήκουν στο $\mathcal{T}_\epsilon^{(n)}$ ονομάζονται ισχυρώς ϵ-τυπικές.

Ισχυρή Τυπικότητα – σχόλια

- Αποδεικνύεται ότι αν μια ακολουθία είναι ισχυρώς τυπική τότε είναι και ασθενώς τυπική.
- Το αντίστροφο δεν ισχύει, όπως είδαμε στο παράδειγμα πηγής Bern$(1/2)$ χωρίς μνήμη.
- Η ισχυρή τυπικότητα είναι πιο ευέλικτη από την ασθενή. Ωστόσο, μπορεί να χρησιμοποιηθεί μόνο για τ.μ. με πεπερασμένο αλφάβητο.
- Μπορούμε να αποδείξουμε τις ίδιες ιδιότητες για τις ισχυρώς τυπικές ακολουθίες όπως και για τις ασθενώς τυπικές με παρόμοιο τρόπο.

Ισχυρή Τυπικότητα – σχόλια (2)

- Συχνά, το ισχυρώς τυπικό σύνολο ορίζεται ως το σύνολο των ακολουθιών που ικανοποιούν τη σχέση

$$|\pi(x|X^n) - p(x)| \leq \epsilon \cdot p(x),$$

για όλα τα $x \in \mathcal{X}$, όπου $\pi(x|X^n) \triangleq \frac{N(x;X^n)}{n}$ είναι ο *τύπος* (type) (ή εμπειρική pmf) της ακολουθίας X^n.

- Παρατηρήστε ότι

$$\sum_{x \in \mathcal{S}_X} |\pi(x|X^n) - p(x)| \leq \sum_{x \in \mathcal{S}_X} \epsilon \cdot p(x) = \epsilon.$$

Επανάληψη Βασικών Ποσοτήτων Θεωρίας Πληροφορίας

Τι εννοούμε με τον όρο "κωδικοποίηση";

- Η αναπαράσταση ενός σήματος/μηνύματος από κάποιο άλλο.
- Μια απεικόνιση από ένα σήμα/μήνυμα σε ένα άλλο.
- Ενδέχεται να μην είναι αντιστρέψιμη (κωδικοποίηση με απώλειες – lossy compression).
- Σε τι χρησιμεύει η κωδικοποίηση;
  1. Συμπίεση (Κωδικοποίηση πηγής)
  2. Μετάδοση μέσω καναλιού (Κωδικοποίηση καναλιού)
  3. Μετατροπή σήματος/μηνύματος σε μορφή την οποία μπορούμε να επεξεργαστούμε. Παράδειγμα: Κβαντισμός συνεχούς σήματος, μετατροπή σήματος σε δυαδική μορφή.
  4. Προστασία δεδομένων και πνευματικής ιδιοκτησίας (Κρυπτογραφία, Υδατογράφηση).

Εντροπία διακριτής τ.μ.

Έστω διακριτή τ.μ. X με συνάρτηση μάζας πιθανότητας (pmf) $p(x)$.

$$H(X) = \mathbb{E}_p\left[\log \frac{1}{p(X)}\right] = \sum_x p(x) \log \frac{1}{p(x)} = -\sum_x p(x) \log p(x).$$

- $\log \frac{1}{p(x)}$: Η πληροφορία που περιέχεται στο ενδεχόμενο $X = x$.
- Η $H(X)$ δεν εξαρτάται από τις τιμές της X, παρά μόνο από την κατανομή της.
- $H(X)$: Το όριο συμπίεσης.
  - Το μέσο μήκος της συντομότερης περιγραφής της X
  - Η μέση πληροφορία που περιέχεται στη X.
  - Η μέση αβεβαιότητα που έχουμε για τη X (πριν μας αποκαλυφθεί η τιμή της).
- Μονάδα μέτρησης: bit ($\log \to \log_2$). Σπανιότερα, nat ($\log \to \ln$).
- $H_b(X) = \log_b a \cdot H_a(X)$.
- Από εδώ και στο εξής $\log$ υπονοεί $\log_2$ (αν και δεν έχει ιδιαίτερη σημασία ποια μονάδα χρησιμοποιούμε).

Από κοινού και υπό συνθήκη εντροπία

- Από κοινού (συνδυασμένη) εντροπία (joint entropy) 2 τ.μ. με από κοινού pmf $p(x, y)$:

$$H(X,Y) = \mathbb{E}_p\left[\log \frac{1}{p(X,Y)}\right]$$
$$= \sum_x \sum_y p(x,y) \log \frac{1}{p(x,y)} = -\sum_x \sum_y p(x,y) \log p(x,y).$$

- Δεσμευμένη εντροπία (conditional entropy) της τ.μ. X δεδομένης της τ.μ. Y:

$$H(X|Y) = \mathbb{E}_p\left[\log \frac{1}{p(X|Y)}\right] = \sum_x \sum_y p(x,y) \log \frac{1}{p(x|y)}$$
$$= -\sum_x \sum_y p(x,y) \log p(x|y) = -\sum_x \sum_y p(y)p(x|y) \log p(x|y)$$
$$= -\sum_y p(y) \sum_x p(x|y) \log p(x|y) = \sum_y p(y) H(X|Y=y).$$

Ιδιότητες Εντροπίας διακριτής τ.μ.

- $H(X) \geq 0$.
- Η εντροπία είναι κοίλη ($\cap$) συνάρτηση της συνάρτησης μάζας πιθανότητας $p(x)$. Θα το αποδείξουμε.
- $H(X) \leq \log |\mathcal{X}|$, όπου $|\mathcal{X}|$ το μέγεθος του αλφαβήτου της X. Το μέγιστο επιτυγχάνεται από την ομοιόμορφη κατανομή: $p(X_i) = \frac{1}{|\mathcal{X}|}$ για όλα τα $X_i \in \mathcal{X}$. Αποδείχθηκε στη "Θεωρία Πληροφορίας".
- $H(X, Y) = H(Y, X)$ (εύκολο, π.χ. με χρήση του ορισμού, δεδομένου ότι $p(x, y) = p(y, x)$).
- Κανόνας αλυσίδας: $H(X_1, X_2, \ldots, X_n) = H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_1, X_2, \ldots, X_{n-1})$. Απόδειξη με χρήση ορισμού και κανόνα Bayes.
- Για ανεξάρτητες τ.μ., $H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i)$.
- Επίσης, εάν οι τ.μ. X και Y είναι ανεξάρτητες, $H(X|Y) = H(X)$ και $H(Y|X) = H(Y)$.
- Γενικά, $H(X|Y) \neq H(Y|X)$.

Ρυθμός Εντροπίας διακριτής πηγής

- Ρυθμός εντροπίας διακριτής πηγής (τυχαίας διαδικασίας):

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) \text{ bits/σύμβολο},$$

εφόσον το όριο συγκλίνει.

- Το όριο συγκλίνει πάντα όταν η πηγή είναι στάσιμη. Στην περίπτωση αυτή, συγκλίνει και η ποσότητα

$$H'(\mathcal{X}) = \lim_{n\to\infty} H(X_n|X_1, X_2, \ldots, X_{n-1})$$

και $H(\mathcal{X}) = H'(\mathcal{X})$.

Ρυθμός Εντροπίας διακριτής πηγής (συνέχεια)

- Εάν οι τ.μ. είναι *ανεξάρτητες*,
  $H(\mathcal{X}) = H'(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i)$.
- Εάν, επιπλέον, οι τ.μ. είναι και ομοίως κατανεμημένες,
  $H(\mathcal{X}) = H'(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} nH(X_i) = H(X_i) = H(X_1)$.
- Για στάσιμες πηγές, ο ρυθμός εντροπίας ποσοτικοποιεί το μέσο ποσό νέας πληροφορίας κάθε φορά που παίρνουμε ένα νέο δείγμα (το ποσό πληροφορίας των innovations για όσους έχουν ασχοληθεί με Θεωρία Εκτίμησης).

Παράδειγμα 2.2 (Cover & Thomas σελ. 74)

- Έστω ακολουθία δυαδικών τ.μ. Bernoulli με $p_i = \Pr\{X_i = 1\}$ που δεν είναι σταθερή, αλλά εξαρτάται από το i ως εξής:

$$p_i = \begin{cases} 0.5 & \text{εάν} \quad 2k < \log\log i \leq 2k+1 \\ 0 & \text{εάν} \quad 2k+1 < \log\log i \leq 2k+2, \end{cases}$$

  για $k = 0, 1, 2, \ldots$.

- Επομένως, κομμάτια όπου $H(X_i) = 1$ ακολουθούνται από εκθετικώς αυξανόμενα κομμάτια όπου $H(X_i) = 0$ κ.ο.κ. Συνεπώς, ο μέσος όρος της $H(X_i)$ μεταβάλλεται συνεχώς και δε συγκλίνει.
- Στη συγκεκριμένη περίπτωση δεν είναι δυνατό να οριστεί ρυθμός εντροπίας $H(\mathcal{X})$.

Σχετική Εντροπία $D(p||q)$

- Η σχετική εντροπία (relative entropy) ή απόσταση Kullback-Leibler μεταξύ δύο κατανομών p και q που ορίζονται στο ίδιο αλφάβητο $\mathcal{A}$ ισούται με

$$D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = \mathbb{E}_p \left[\log \frac{p(X)}{q(X)} \right].$$

- Προσοχή: Η μέση τιμή είναι ως προς την κατανομή p.
- Από πού πηγάζει αυτός ο ορισμός; Όπως είδαμε στη "Θεωρία Πληροφορίας", η $D(p||q)$ ποσοτικοποιεί τα επιπλέον bits που χρειαζόμαστε για να συμπιέσουμε μια τ.μ. με πραγματική κατανομή p όταν για τη συμπίεση χρησιμοποιείται η κατανομή q.

Σχετική Εντροπία $D(p||q)$ (συνέχεια)

- Όταν χρησιμοποιείται κώδικας Shannon, $H(X)+D(p||q) \leq \mathbb{E}[l^*] < H(X) + D(p||q) + 1$, όπου $\mathbb{E}[l^*]$ είναι το μέσο μήκος του κώδικα Shannon για την κατανομή q, ενώ η πραγματική κατανομή της X είναι η p.
- $D(p||q) \geq 0$. Αποδείχθηκε στη "Θεωρία Πληροφορίας" με χρήση της ανισότητας Jensen και του γεγονότος ότι η $\log$ είναι κοίλη ($\cap$). Θα επαναλάβουμε την απόδειξη στο μάθημα.
- Ωστόσο, η $D(p||q)$ δεν είναι απόσταση κατά την αυστηρή έννοια:
  - $D(p||q) \neq D(q||p)$.
  - Επίσης, δεν ισχύει η τριγωνική ανισότητα.

Δεσμευμένη Σχετική Εντροπία και Κανόνας Αλυσίδας

- Δεσμευμένη σχετική εντροπία (conditional relative entropy):

$$D(p(y|x)||q(y|x)) = \mathbb{E}_p\left[\log \frac{p(Y|X)}{q(Y|X)}\right] = \sum_x \sum_y p(x,y) \log \frac{p(y|x)}{q(y|x)}.$$

  - Προσοχή: Μέση τιμή ως προς την $p(x,y)$.
- Κανόνας αλυσίδας για τη σχετική εντροπία

$$D(p(x,y)||q(x,y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x)).$$

  - Απόδειξη: Απλή, με χρήση ορισμού (Cover & Thomas Theorem 2.5.3).

Αμοιβαία Πληροφορία $I(X;Y)$

- Έστω μια τ.μ. $X \sim p(X)$. Εάν μας γνωστοποιηθεί η τιμή της τ.μ. Y, η κατανομή πιθανότητας της X αλλάζει σε $p(X|Y)$. Επομένως, κατά μέσο όρο, γνώση της Y αλλάζει την αβεβαιότητα που έχουμε για τη X κατά $\mathbb{E}_p\left[\frac{p(X|Y)}{p(X)}\right]$ (η μέση τιμή υπολογίζεται για όλες τις τιμές των X και Y).

- Συνεπώς,

$$I(X;Y) \triangleq \mathbb{E}_p\left[\log\frac{p(X|Y)}{p(X)}\right] = \sum_x\sum_y p(x,y)\log\frac{p(x|y)}{p(x)}$$

$$= \sum_x\sum_y p(x,y)\log\frac{p(x|y)p(y)}{p(x)p(y)} = \sum_x\sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

$$= D(p(x,y)||p(x)p(y)) = \mathbb{E}_p\left[\log\frac{p(X,Y)}{p(X)p(Y)}\right].$$

Αμοιβαία Πληροφορία $I(X;Y)$ (2)

- Προφανώς (από την προηγούμενη σχέση), $I(X;Y) = I(Y;X)$. Άρα, αποκάλυψη της X οδηγεί στην ίδια μεταβολή της αβεβαιότητας για την Y κατά μέσο όρο.
- Η ποσότητα $I(X;Y)$ ονομάζεται αμοιβαία πληροφορία. Έχουμε δει (και θα το αποδείξουμε, και πάλι, αργότερα) ότι $I(X;Y) \geq 0$. Επομένως, αποκάλυψη της τιμής της Y ελαττώνει την αβεβαιότητα για τη X κατά μέσο όρο.
- Προσοχή: Για κάποιες τιμές της Y, ενδέχεται $I(X;Y=y) < 0$. Ωστόσο, ισχύει πάντα $I(X;Y) = \mathbb{E}_{p_Y}[I(X;Y=y)] \geq 0$.

Αμοιβαία Πληροφορία $I(X;Y)$ (3)

- Μια διαφορετική ερμηνεία της αμοιβαίας πληροφορίας με βάση τη σχετική εντροπία: Η πληροφορία που "χάνουμε" εάν θεωρήσουμε ότι οι X και Y είναι ανεξάρτητες, ενώ, στην πραγματικότητα, δεν είναι.
- $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$. Προκύπτει από τον ορισμό (αποδείχτηκε στη "Θεωρία Πληροφορίας").

Αμοιβαία Πληροφορία $I(X;Y)$ (4)

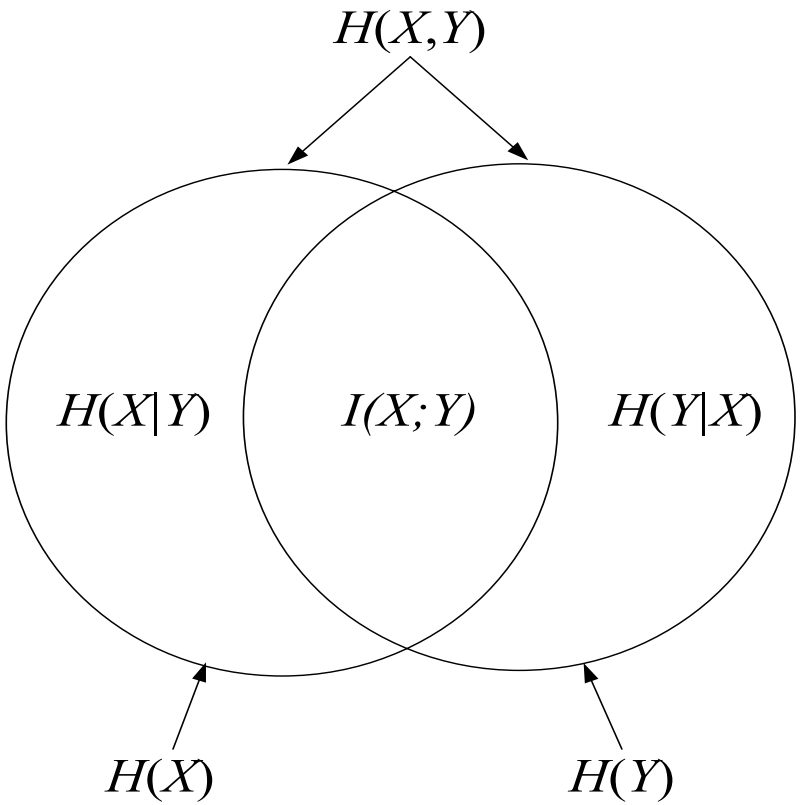
- $I(X;X) = H(X) - H(X|X) = H(X)$. Η X περιέχει όλη την πληροφορία για τον εαυτό της.
- Κανόνας αλυσίδας για την αμοιβαία πληροφορία:

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

- Απόδειξη: Εύκολα, από κανόνα αλυσίδας εντροπίας και χρήση $I(X_1, X_2, \ldots, X_n; Y) = H(X_1, X_2, \ldots, X_n) - H(X_1, X_2, \ldots, X_n|Y)$.
- Υπό συνθήκη αμοιβαία πληροφορία: $I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$.

Διάγραμμα Venn

Η σχέση μεταξύ εντροπίας, δεσμευμένης εντροπίας και αμοιβαίας πληροφορίας μπορεί να αναπαρασταθεί και με χρήση διαγράμματος Venn.



Ιδιότητες των βασικών μεγεθών της Θεωρίας Πληροφορίας

Κυρτές (convex) και κοίλες (concave) συναρτήσεις

- Ορισμός 2.2. Μια συνάρτηση $f(x)$ είναι κυρτή ($\cup$) σε διάστημα (a, b) εάν, για κάθε $x_1, x_2 \in (a, b)$ και $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

- Έχουμε χρησιμοποιήσει εμμέσως το γεγονός ότι το διάστημα (a, b) είναι κυρτό: $\forall x_1, x_2 \in (a, b)$, $\lambda x_1 + (1-\lambda)x_2 \in (a, b)$ για $0 \leq \lambda \leq 1$.

- Ορισμός 2.3. Μια συνάρτηση $f(x)$ είναι αυστηρώς κυρτή (strictly convex) εάν η ισότητα στην παραπάνω σχέση ισχύει μόνο για $\lambda = 0$ ή $\lambda = 1$.

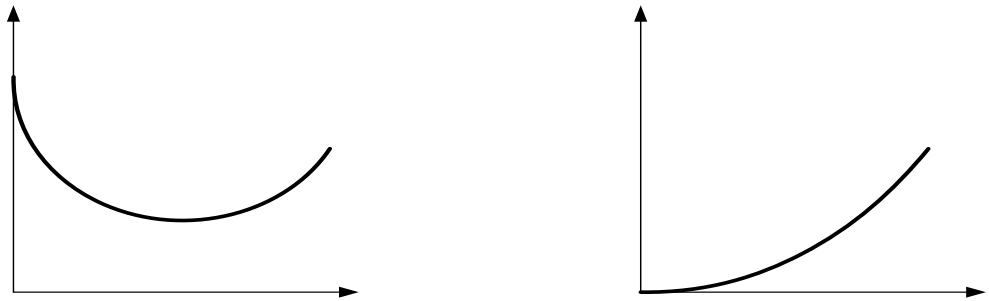
- Πρακτικά, μια συνάρτηση είναι κυρτή όταν μια χορδή που ενώνει δύο οποιεσδήποτε τιμές της δε βρίσκεται ποτέ "κάτω" από τη συνάρτηση.

- Παραδείγματα κυρτών συναρτήσεων: x^2, $|x|$, e^x, $x \log x$ (για $x \geq 0$).

Κυρτές (convex) και κοίλες (concave) συναρτήσεις (συνέχεια)

- Ορισμός 2.4. Μια συνάρτηση $f(x)$ είναι (αυστηρώς) κοίλη ($\cap$) σε διάστημα (a, b) εάν η $-f(x)$ είναι (αυστηρώς) κυρτή.
  - Παραδείγματα κοίλων συναρτήσεων: $\log x$, $\sqrt{x}$ (για $x \geq 0$).
  - Η συνάρτηση $ax + b$ (affine) είναι κυρτή και κοίλη.

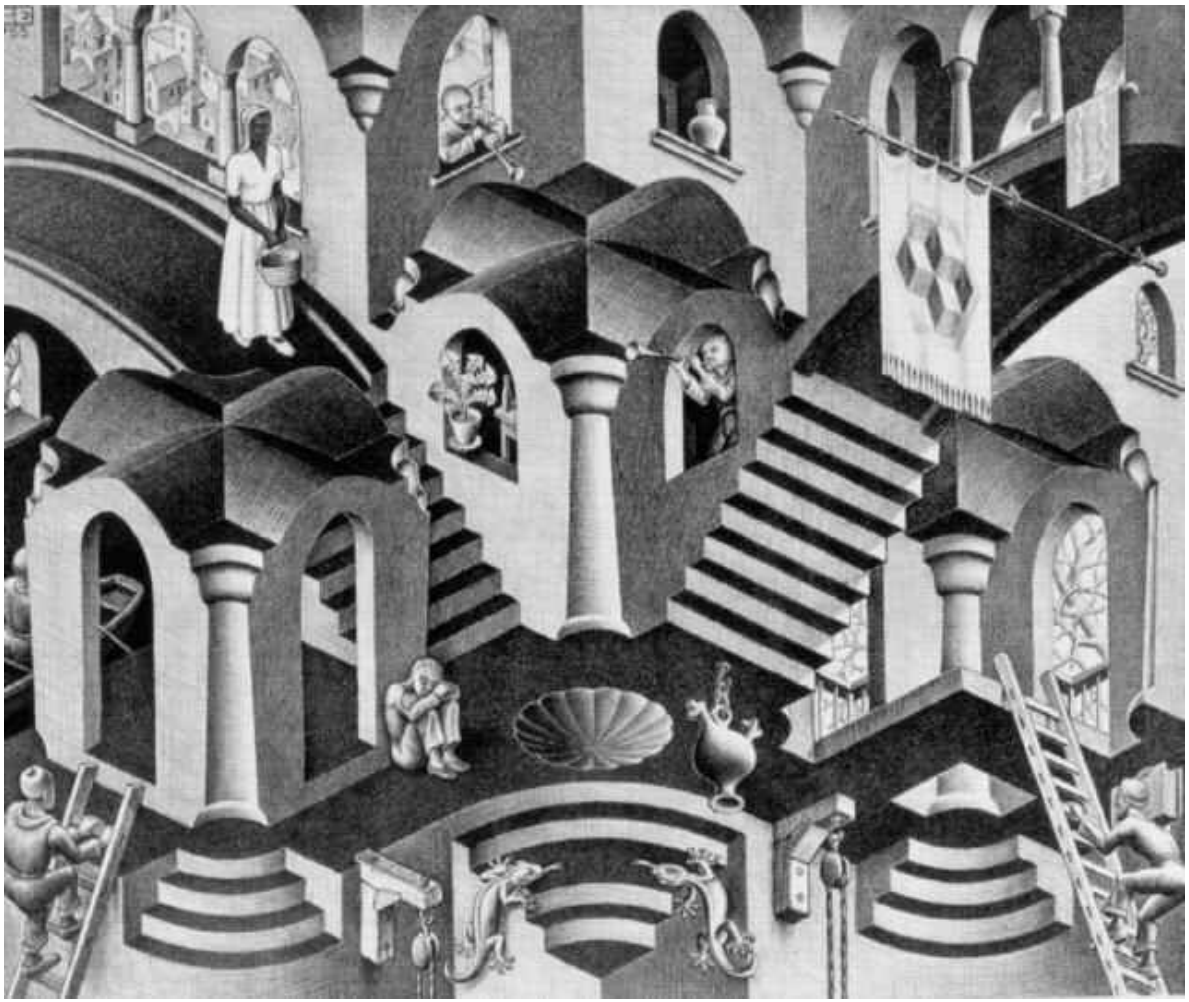
Παραδείγματα κυρτών και κοίλων συναρτήσεων



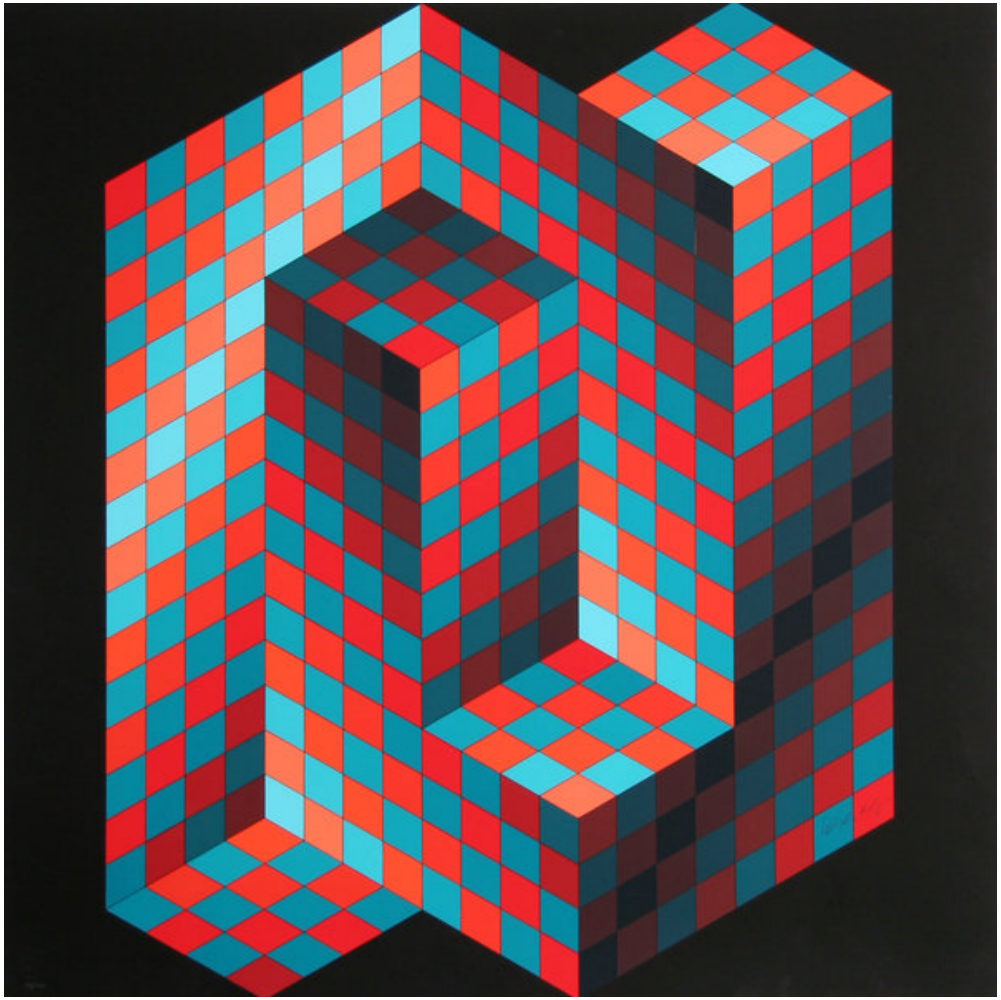
(α) Κυρτές συναρτήσεις

(β) Κοίλες συναρτήσεις

M. C. Escher, Convex and Concave, 1955



V. Vasarely, Gestalt 4, 1970



Ανισότητα Jensen

- Θεώρημα 2.5. Μια *διαφορίσιμη* συνάρτηση είναι (αυστηρώς) κυρτή ($\cup$) σε ένα διάστημα όταν έχει μη αρνητική (θετική) δεύτερη παράγωγο στο διάστημα αυτό.
  - Απόδειξη: Σε βιβλία ανάλυσης ή Cover & Thomas Theorem 2.6.1
- Θεώρημα 2.6. (Ανισότητα Jensen): Εάν η συνάρτηση f είναι κυρτή και η X είναι τυχαία μεταβλητή,

Ανισότητα Jensen

$$\mathbb{E}\left[f(X)\right] \geq f\left(\mathbb{E}[X]\right)$$

  - Απόδειξη: με επαγωγή (induction) για διακριτές τ.μ. (Cover & Thomas)

Απόδειξη ανισότητας Jensen

- Για τ.μ. με δύο ενδεχόμενα, από τον ορισμό της κυρτότητας, $p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 x_1 + p_2 x_2)$ (δεδομένου ότι $p_2 = 1 - p_1$).
- Έστω ότι η σχέση ισχύει για τ.μ. με $k-1$ ενδεχόμενα.
- Θέτουμε $p'_i = \frac{p_i}{1-p_k}$, για $i = 1, 2, \ldots, k-1$.

$$\sum_{i=1}^{k} p_i f(x_i) = p_k f(x_k) + (1-p_k) \sum_{i=1}^{k-1} p'_i f(x_i)$$

$$\overset{(a)}{\geq} p_k f(x_k) + (1-p_k) f\left(\sum_{i=1}^{k-1} p'_i x_i \right)$$

$$\overset{(b)}{\geq} f\left(p_k x_k + (1-p_k) \sum_{i=1}^{k-1} p'_i x_i \right) = f\left(\sum_{i=1}^{k} p_i x_i \right),$$

όπου στο (a) χρησιμοποιήθηκε η παραδοχή ότι η ανισότητα Jensen ισχύει για $k-1$, ενώ στο (b) χρησιμοποιήθηκε το γεγονός ότι η ανισότητα ισχύει για $k = 2$.

Ανισότητα πληροφορίας (ή Gibbs): $D(p||q) \geq 0$

- $D(p||q) = 0 \Leftrightarrow p(x) = q(x)$ για κάθε $x \in \mathcal{X}$.
- Απόδειξη με χρήση ορισμού και ανισότητας Jensen:
  Έστω $\mathcal{A} = \{x : p(x) > 0\}$.

$$-D(p||q) = -\sum_{x\in\mathcal{A}} p(x) \log \frac{p(x)}{q(x)} = \sum_{x\in\mathcal{A}} p(x) \log \frac{q(x)}{p(x)} =$$

$$\overset{(a)}{\leq} \log \sum_{x\in\mathcal{A}} p(x) \frac{q(x)}{p(x)} = \log \sum_{x\in\mathcal{A}} q(x) \overset{(b)}{\leq} \log \sum_{x\in\mathcal{X}} q(x) = \log 1 = 0.$$

- Στο (a) χρησιμοποιήθηκε το γεγονός ότι η $\log t$ είναι αυστηρώς κοίλη συνάρτηση του t. (b) γιατί;
- Η ισότητα ισχύει εάν και μόνο εάν $q(x)/p(x) = c$ για όλα τα x, δηλαδή εάν $q(x) = cp(x)$. Επίσης, πρέπει $\sum_{x\in\mathcal{A}} q(x) = \sum_{x\in\mathcal{X}} q(x) = \sum_{x\in\mathcal{X}} cp(x) = c = 1$. Συνεπώς, $D(p||q) = 0 \Leftrightarrow p(x) = q(x)$ για όλα τα $x \in \mathcal{A}$.

Συνέπειες ανισότητας πληροφορίας

- Η αμοιβαία πληροφορία είναι πάντοτε μη αρνητική: Για οποιεσδήποτε τ.μ. X και Y, $\boxed{I(X;Y) \geq 0}$. Προκύπτει άμεσα από τον ορισμό της $I(X;Y)$ και από την ανισότητα πληροφορίας.
- $D(p(y|x)||q(y|x)) \geq 0$ (Γιατί; Πότε ισχύει η ισότητα;)
- $I(X;Y|Z) \geq 0$.
- $H(X|Y) \leq H(X)$.
  Δεδομένου ότι $I(X;Y) \geq 0 \Rightarrow H(X) - H(X|Y) \geq 0$.
- Προσοχή: Δεν ισχύει πάντα $H(X|Y=y) \leq H(X)$
  (και, επομένως, δεν ισχύει πάντα ότι $I(X;Y=y) \geq 0$).

Σχέση μεταξύ $I(X;Y)$ και $I(X;Y|Z)$

- Σε αντίθεση με την υπό συνθήκη εντροπία (για την οποία ισχύει $H(X|Z) \leq H(X)$), δεν υπάρχει κάποια γενική ανισότητα που συνδέει την $I(X;Y)$ και την $I(X;Y|Z)$.
- Δύο σημαντικές ειδικές περιπτώσεις
  - Εάν $p(x,y,z) = p(x)p(z)p(y|x,z)$, $I(X;Y|Z) \geq I(X;Y)$. Θα το αποδείξουμε σύντομα όταν θα μιλήσουμε για την κυρτότητα της $I(X;Y)$.
  - Εάν οι X, Y και Z σχηματίζουν ακολουθία Markov (δηλαδή $X \to Y \to Z$), $I(X;Y|Z) \leq I(X;Y)$. Θα το αποδείξουμε σύντομα όταν αναφερθούμε στην Ανισότητα Επεξεργασίας Δεδομένων.

Φράγμα Ανεξαρτησίας (Independence Bound) Από Κοινού Εντροπίας

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

- Η ισότητα ισχύει εάν και μόνο εάν οι X_i είναι ανεξάρτητες.

Άνω φράγμα $H(X)$ δεδομένου του πλήθους ενδεχομένων $|\mathcal{X}|$

- Θεώρημα 2.7. $H(X) \leq \log |\mathcal{X}|$, όπου $|\mathcal{X}|$ ο αριθμός των στοιχείων (cardinality) του $\mathcal{X}$. Η ισότητα ισχύει εάν και μόνο εάν η X είναι ομοιόμορφα κατανεμημένη στο $\mathcal{X}$.
- Απόδειξη
  - Έστω $u(x) = \frac{1}{|\mathcal{X}|}$ η (διακριτή) ομοιόμορφη κατανομή μάζας πιθανότητας στο σύνολο $\mathcal{X}$ και $p(x)$ η κατανομή μάζας πιθανότητας της X. Από τον ορισμό της σχετικής εντροπίας, $D(p||u) = \sum p(x) \log \frac{p(x)}{u(x)} = \log |\mathcal{X}| - H(X)$.
  - Από την ανισότητα πληροφορίας, $0 \leq D(p||u) = \log |\mathcal{X}| - H(X) \Rightarrow H(X) \leq \log |\mathcal{X}|$.
  - Η ισότητα ισχύει εάν $D(p||u) = 0$, δηλαδή εάν και μόνο εάν $p(x) = u(x)$.

Ανισότητα log sum

- Ανισότητα log sum: Για *μη αρνητικούς αριθμούς* $a_1, a_2, \ldots, a_n$ και $b_1, b_2, \ldots, b_n$,

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}.$$

Η ισότητα ισχύει εάν και μόνο εάν $\frac{a_i}{b_i} = c$, όπου c σταθερά.

Απόδειξη ανισότητας log sum

- Απόδειξη: Έστω ότι $a_i > 0$ και $b_i > 0$ (αποδείξτε ως άσκηση την περίπτωση που δεν υπάρχει i για το οποίο να ισχύει $a_i b_i > 0$). Η συνάρτηση $t \log t$ είναι αυστηρώς κυρτή ($\cup$) ($(t \log t)'' = \frac{1}{t} \log e > 0$ για θετικό t). Από την ανισότητα Jensen,

$$\sum \lambda_i f(t_i) \geq f\left(\sum \lambda_i t_i\right),$$

για $\lambda_i \geq 0, \sum_i \lambda_i = 1$. Θέτοντας $\lambda_i = \frac{b_i}{\sum_{j=1}^{n} b_j}$ και $t_j = \frac{a_i}{b_i}$,

$$\sum \frac{a_i}{\sum b_j} \log \frac{a_i}{b_i} \geq \sum \frac{a_i}{\sum b_j} \log \sum \frac{a_i}{\sum b_j} \Rightarrow$$

$$\sum a_i \log \frac{a_i}{b_i} \geq \left(\sum a_i\right) \log \frac{\sum a_i}{\sum b_i}.$$

Η $D(p||q)$ είναι κυρτή ($\cup$)

- Θεώρημα 2.8. Η $D(p||q)$ είναι κυρτή στο ζεύγος κατανομών (p, q). Δηλαδή, εάν (p_1, q_1) και (p_2, q_2) είναι ζεύγη συναρτήσεων μάζας πιθανότητας,

$$D(\lambda p_1 + (1-\lambda)p_2||\lambda q_1 + (1-\lambda)q_2) \leq \lambda D(p_1||q_1) + (1-\lambda)D(p_2||q_2),$$

για $0 \leq \lambda \leq 1$.

- Απόδειξη: Με χρήση της ανισότητας log sum. Για οποιοδήποτε ενδεχόμενο x,

$$(\lambda p_1(x) + (1-\lambda)p_2(x)) \log \frac{\lambda p_1(x) + (1-\lambda)p_2(x)}{\lambda q_1(x) + (1-\lambda)q_2(x)} \leq$$
$$\lambda p_1(x) \log \frac{\lambda p_1(x)}{\lambda q_1(x)} + (1-\lambda)p_2(x) \log \frac{(1-\lambda)p_2(x)}{(1-\lambda)q_2(x)}.$$

Αθροίζοντας για όλα τα ενδεχόμενα x και με χρήση του ορισμού της σχετικής εντροπίας προκύπτει η κυρτότητα της D.

Η εντροπία είναι κοίλη ($\cap$)

- Είδαμε ότι, εάν $u(x)$ είναι η ομοιόμορφη διακριτή κατανομή, $D(p||u) = \sum p(x) \log \frac{p(x)}{u(x)} = \log |\mathcal{X}| - H(X) \Rightarrow H(X) = \log |\mathcal{X}| - D(p||u)$.
- Δεδομένου ότι η $D(p||u)$ είναι κυρτή, η $-D(p||u)$ (και, επομένως, και η εντροπία) είναι κοίλη.

Θεώρημα 2.9. Συνεπώς, για την εντροπία ισχύει $H(\lambda p_1 + (1-\lambda)p_2) \geq \lambda H(p_1) + (1-\lambda)H(p_2)$.

- Για εναλλακτική απόδειξη, χωρίς χρήση ανισότητας log sum δείτε Cover & Thomas Theorem 2.7.3.

Η $I(X;Y)$ είναι κοίλη ($\cap$) συνάρτηση της $p(x)$ για δεδομένη $p(y|x)$

- Απόδειξη:
  - $I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p(x)H(Y|X=x)$.
  - 1ος όρος: $p(y) = \sum_x p(y|x)p(x)$. Συνεπώς, για δεδομένη $p(y|x)$, η $p(y)$ είναι γραμμική συνάρτηση της $p(x)$. Η $H(Y)$ είναι κοίλη συνάρτηση της $p(y)$ και, επομένως, και της $p(x)$.
  - 2ος όρος: Γραμμική συνάρτηση της $p(x)$.
  - Επομένως, η $I(X;Y)$ είναι κοίλη ($\cap$) συνάρτηση της $p(x)$ για δεδομένη $p(y|x)$.
  - Θυμηθείτε ότι, σε διακριτά κανάλια χωρίς μνήμη, η χωρητικότητα ισούται με τη μέγιστη τιμή της $I(X;Y)$. Το γεγονός ότι η $I(X;Y)$ είναι κοίλη για δεδομένο κανάλι ($p(y|x)$) σημαίνει ότι, εάν βρούμε ένα τοπικό μέγιστο, τότε είναι και ολικό μέγιστο και η κατανομή (ή οι κατανομές) πηγής $p^*(x)$ που μεγιστοποιεί(ούν) την $I(X;Y)$ είναι αυτή(ές) η(οι) οποία(ες) επιτυγχάνει(ουν) τη χωρητικότητα.

Η $I(X;Y)$ είναι κυρτή ($\cup$) συνάρτηση της $p(y|x)$ για δεδομένη $p(x)$

- Απόδειξη:
  - Έστω δύο υπό συνθήκη κατανομές μάζας πιθανότητας $p_1(y|x)$ και $p_2(y|x)$. $p_1(x,y) = p(x)p_1(y|x)$ και $p_2(x,y) = p(x)p_2(y|x)$. Επίσης, $p_1(y) = \sum_x p_1(x,y)$ και $p_2(y) = \sum_x p_2(x,y)$. Η περιθώρια κατανομή των $p_1(x,y)$ και $p_2(x,y)$ ως προς x είναι η $p(x)$.
  - Έστω, τώρα, η υπό συνθήκη κατανομή που προκύπτει από την "ανάμιξη" των $p_1(y|x)$ και $p_2(y|x)$:

$$p_\lambda(y|x) = \lambda p_1(y|x) + (1-\lambda)p_2(y|x),\ 0 \le \lambda \le 1.$$

Συνεπώς, ισχύει, επίσης,

$$\begin{aligned} p_\lambda(x,y) &= p_\lambda(y|x)p(x) = \lambda p_1(y|x)p(x) + (1-\lambda)p_2(y|x)p(x) \\ &= \lambda p_1(x,y) + (1-\lambda)p_2(x,y) \end{aligned}$$

και

$$p_\lambda(y) = \lambda p_1(y) + (1-\lambda)p_2(y).$$

Η $I(X;Y)$ είναι κυρτή ($\cup$) συνάρτηση της $p(y|x)$ για δεδομένη $p(x)$ (2)

- Απόδειξη (συνέχεια):
  - Ορίζουμε την κατανομή $q_\lambda(x,y)$ ως το γινόμενο των περιθώριων κατανομών:

$$q_\lambda(x,y) = p(x)p_\lambda(y) = \lambda q_1(x,y) + (1-\lambda)q_2(x,y).$$

  - Από τον ορισμό της αμοιβαίας πληροφορίας παρατηρούμε ότι

$$\begin{aligned} I(X;Y) &= D(p_\lambda(x,y)||p_\lambda(x)p_\lambda(y)) = D(p_\lambda(x,y)||p(x)p_\lambda(y)) \\ &= D(p_\lambda(x,y)||q_\lambda(x,y)). \end{aligned}$$

  - Η $D(p||q)$ είναι κυρτή συνάρτηση του ζεύγους (p,q). Επομένως, και η $I(X;Y)$ είναι κυρτή συνάρτηση της $p(y|x)$.

Η $I(X;Y)$ είναι κυρτή ($\cup$) συνάρτηση της $p(y|x)$ για δεδομένη $p(x)$ (3)

- Συνεπώς, για δεδομένη κατανομή πηγής, υπάρχει κάποιο κανάλι το οποίο ελαχιστοποιεί την πληροφορία που μπορούμε να μεταδώσουμε στο δέκτη.
- Επίσης, για δεδομένη κατανομή εισόδου, $p(x)$, η αμοιβαία πληροφορία όταν χρησιμοποιούμε κανάλι που προκύπτει από το "μέσο όρο" δύο καναλιών δεν μπορεί να υπερβεί το μέσο όρο των αμοιβαίων πληροφοριών για κάθε κανάλι ξεχωριστά (που αντιστοιχούν στην ίδια $p(x)$).
  - Αυτό έχει ως αποτέλεσμα η χωρητικότητα ενός καναλιού που μεταβάλλεται να είναι μεγαλύτερη όταν ο πομπός γνωρίζει το κανάλι και μπορεί να προσαρμόζει τις κωδικές λέξεις και το ρυθμό μετάδοσης.

Παράδειγμα 2.3

- Υποθέτουμε ότι $p(x, y, z) = p(x)p(z)p(y|x, z)$.
- Θα αποδείξουμε ότι $I(X; Y|Z) = I(X; Y)$.
- Από τον ορισμό της $I(X; Y|Z)$,

$$
\begin{aligned}
I(X; Y|Z) &= \sum_x \sum_y \sum_z p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} \\
&= \sum_x \sum_y \sum_z p(x)p(y|x, z)p(z) \log \frac{p(y|x, z)}{p(y|z)} \\
&= \sum_z p(z) \sum_x \sum_y p(x)p(y|x, z) \log \frac{p(y|x, z)}{p(y|z)}
\end{aligned}
$$

Παράδειγμα 2.3 (2)

$$I(X;Y|Z) = \sum_z p(z) \sum_x \sum_y p(x)p(y|x,z) \log \frac{p(y|x,z)}{p(y|z)}.$$

- Μπορούμε να δούμε τις $p(y|x,z)$ ως μια οικογένεια κατανομών $p^{(z)}(y|x)$ με δείκτη Z. Δηλαδή, σε κάθε τιμή z της Z αντιστοιχεί μία κατανομή $p^{(z)}(y|x) \triangleq p(y|x,z)$.
- Αποδείξαμε, όμως, ότι, για δεδομένη $p(x)$, η $I(X;Y)$ είναι κυρτή ($\cup$) συνάρτηση της $p(y|x)$.

Παράδειγμα 2.3 (3)

- Επομένως, από την Ανισότητα Jensen,

$$
\begin{aligned}
I(X;Y|Z) &= \sum_z p(z) \sum_x \sum_y p(x)p(y|x,z) \log \frac{p(y|x,z)}{p(y|z)} \\
&\geq \sum_x \sum_y p(x) \left\{ \sum_z p(z)p(y|x,z) \right\} \log \frac{\sum_z p(z)p(y|x,z)}{\sum_z p(z)p(y|z)} \\
&= \sum_x \sum_y p(x)p(y|x) \log \frac{p(y|x)}{p(y)} = I(X;Y).
\end{aligned}
$$