EE728
Προχωρημένα Θέματα Θεωρίας Πληροφορίας
2η διάλεξη

Δημήτρης-Αλέξανδρος Τουμπακάρης

Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών

2 Μαρτίου 2010

Περιεχόμενα σημερινού μαθήματος

Αντιστοιχία με βιβλία Cover & Thomas και Gallager

- Cover & Thomas: 2.3 – 2.7, 3.1, 3.2
- Gallager: 2.2, 2.3

Ρυθμός Εντροπίας διακριτής πηγής

Ορισμός 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 2.1 (Cover 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 $D(p||q)$

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

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

- 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)

- $H(X) + D(p||q) \leq E[l^*] < H(X) + D(p||q) + 1$, όπου $E[l^*]$ είναι το μέσο μήκος του βέλτιστου κώδικα πηγής 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):

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

- 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 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 $E_p\left[\frac{p(X|Y)}{p(X)}\right]$ (η μέση τιμή υπολογίζεται για όλες τις τιμές των X και Y).

Ορισμός Συνεπώς,

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

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

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

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 $I(X;Y) = E_{p(Y)}[I(X;Y=y)] \geq 0$.

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 (αποδείχθηκε στη "Θεωρία Πληροφορίας").

Αμοιβαία Πληροφορία $I(X;Y)$ (4)

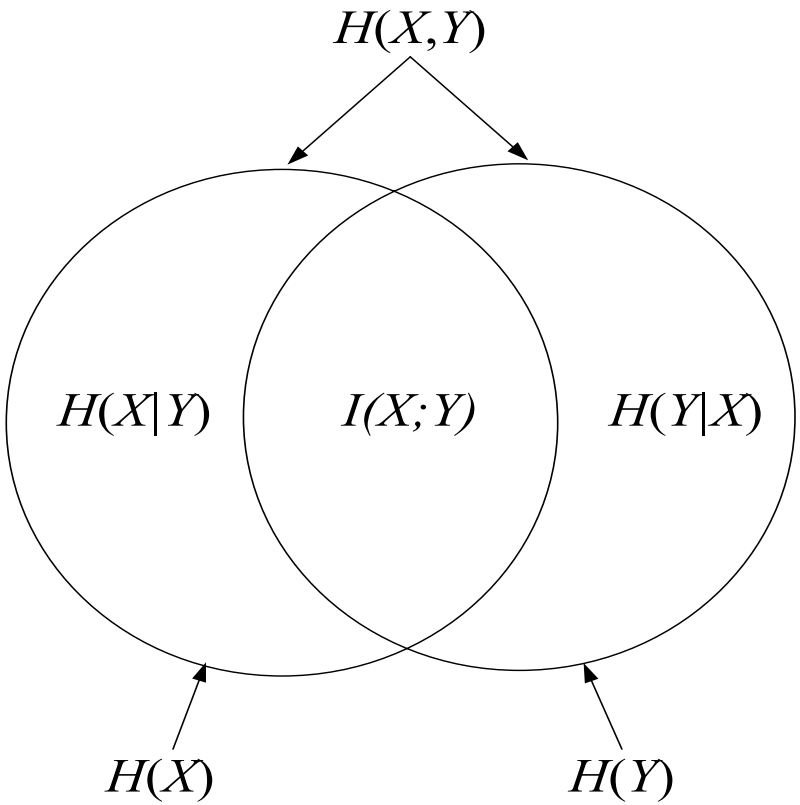
- $I(X;X) = H(X) - H(X|X) = H(X)$. Η X περιέχει όλη την πληροφορία για τον εαυτό της.
- Κανόνας αλυσίδας για την αμοιβαία πληροφορία:

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

- 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.



Κυρτές (convex) και κοίλες (concave) συναρτήσεις

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

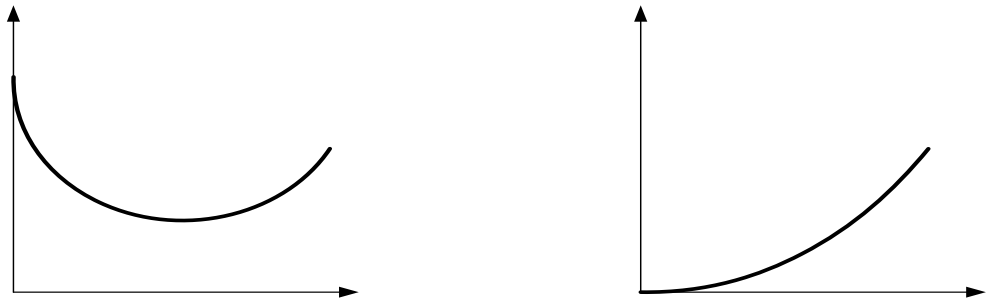
- Πρακτικά, μια συνάρτηση είναι κυρτή όταν μια χορδή που ενώνει δύο οποιεσδήποτε τιμές της δε βρίσκεται ποτέ "κάτω" από τη συνάρτηση.
- Παραδείγματα κυρτών συναρτήσεων: x^2, $|x|$, e^x, $x \log x$ (για $x \geq 0$).

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

Ορισμός 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.

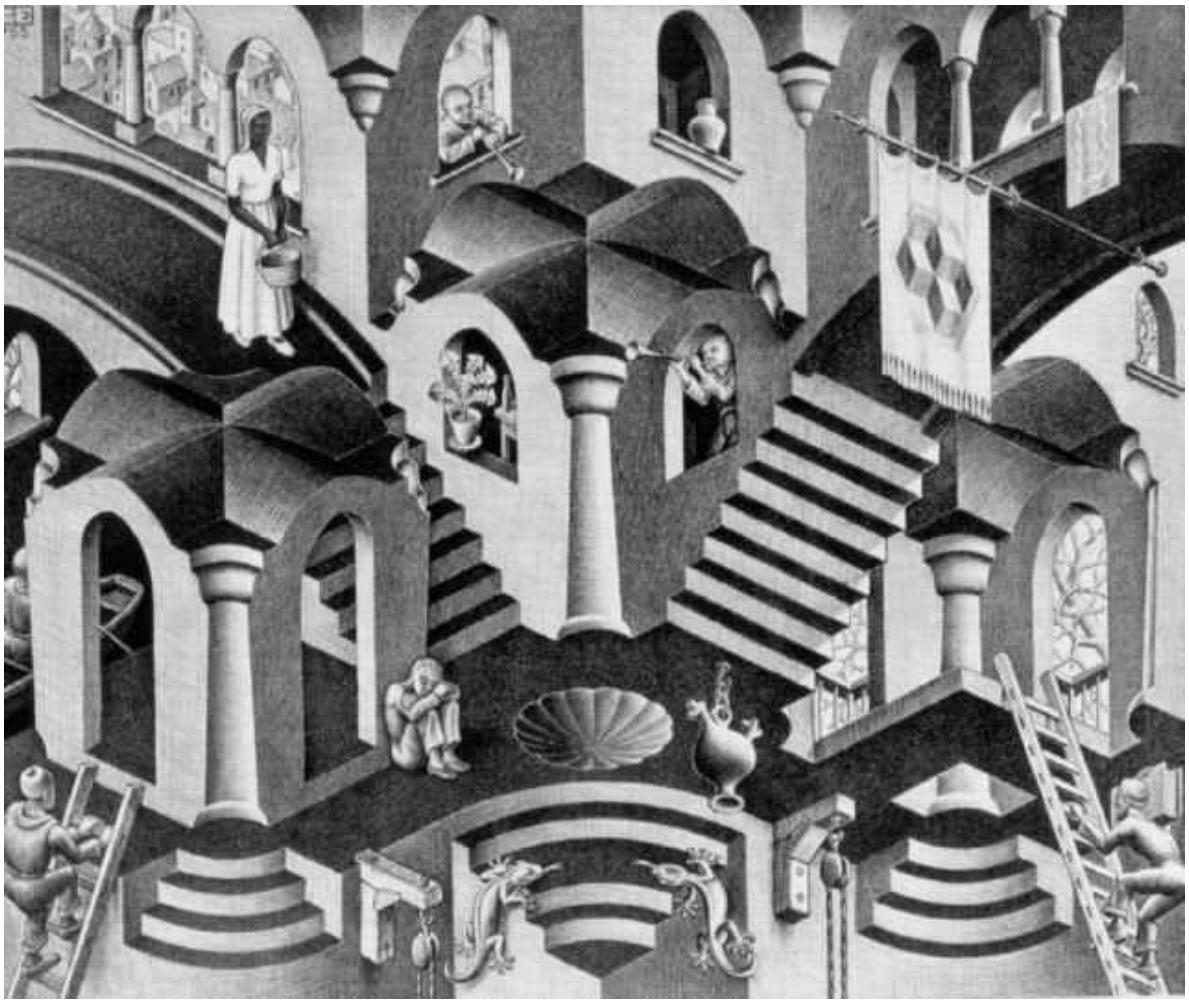
Παραδείγματα κυρτών και κοίλων συναρτήσεων



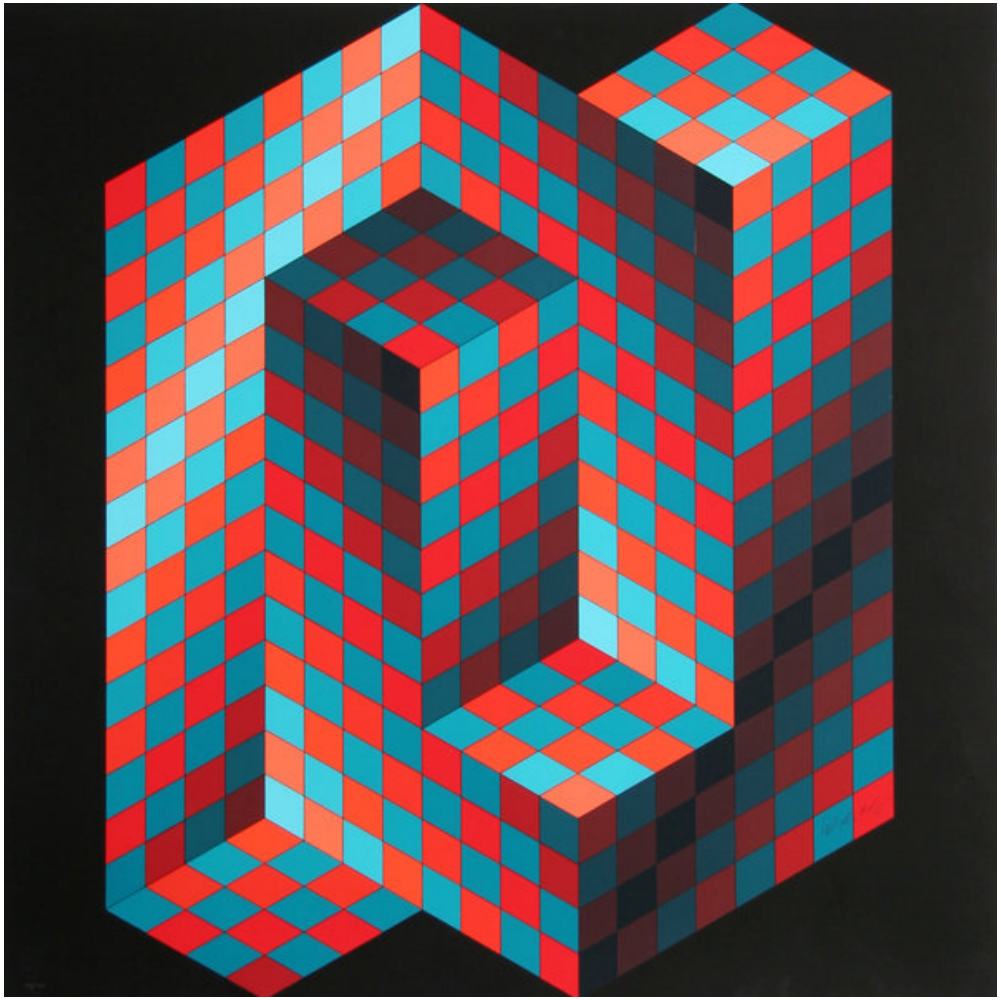
(α) Κυρτές συναρτήσεις

(β) Κοίλες συναρτήσεις

M. C. Escher, Convex and Concave, 1955



V. Vasarely, Gestalt 4, 1970



Ανισότητα Jensen

Θεώρημα 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 Theorem 2.6.1

Θεώρημα Ανισότητα Jensen: Εάν η συνάρτηση f είναι κυρτή και η X είναι τυχαία μεταβλητή,

$$\boxed{Ef(X) \geq f(EX)}$$

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

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 $I(X;Y=y) \geq 0$).

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 $H(X) \leq \log|\mathcal{X}|$, όπου $|\mathcal{X}|$ ο αριθμός στοιχείων (cardinality) του $\mathcal{X}$. Η ισότητα ισχύει εάν και μόνο εάν η X είναι ομοιόμορφα κατανεμημένη στο $\mathcal{X}$.

- 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 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 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 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 $p(x)$

- 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)

- 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 $p(x)$).

Η Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP)

Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP) – Εισαγωγή

- Θεωρούμε μια ακολουθία ανεξάρτητων, ομοίως κατανεμημένων (i.i.d.) διακριτών τ.μ. X_i: $X_1^n = X_1, X_2, \ldots, X_n$.
- Η από κοινού συνάρτηση μάζας πιθανότητας των τ.μ. που αποτελούν την ακολουθία ισούται με $p(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} p(X_i)$.

Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP) – Εισαγωγή (2)

- Asymptotic Equipartition Property -- AEP:
  Αυξάνοντας το μήκος της ακολουθίας,

$$
\begin{aligned}
-\frac{1}{n}\lim_{n\to\infty}\log p(X_1, X_2, \ldots, X_n) &= -\lim_{n\to\infty}\frac{1}{n}\log\prod_{i=1}^{n} p(X_i) \\
&= -\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log p(X_i) \\
&= -\mathbb{E}[\log p(X)] = H(X),
\end{aligned}
$$

  από τον Ασθενή Νόμο Μεγάλων Αριθμών (Weak Law of Large Numbers).

Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP) – Εισαγωγή (3)

- Επομένως, εάν σχηματίσουμε μια ακολουθία πολύ μεγάλου μήκους, η από κοινού συνάρτηση κατανομής μάζας πιθανότητας θα συγκλίνει κατά πιθανότητα στην τιμή $2^{-nH(X)}$.
- Θα αποδείξουμε ότι υπάρχουν περίπου $2^{nH(X)}$ τέτοιες, τυπικές, ακολουθίες και ότι το άθροισμα των από κοινού συναρτήσεων μάζας πιθανότητάς τους προσεγγίζει το 1.
- Το άθροισμα των πιθανοτήτων των υπόλοιπων, μη τυπικών, ακολουθιών μήκους n τείνει στο 0.
- Επομένως, μπορούμε να κωδικοποιήσουμε μόνο τις τυπικές ακολουθίες $\rightarrow$ χρειαζόμαστε nH bits αντί για n.

Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP) – Εισαγωγή (4)

- Επειδή η πιθανότητα να εμφανιστεί μη τυπική ακολουθία τείνει στο 0, η πιθανότητα να μην μπορούμε να κωδικοποιήσουμε την ακολουθία X_1^n με χρήση nH bits τείνει στο 0 για $n \to \infty$.
- Το AEP αποτελεί στυλοβάτη της Θεωρίας Πληροφορίας.

Είδη σύγκλισης (υπενθύμιση)

Μια ακολουθία τ.μ. $X_1, X_2, \ldots$ συγκλίνει σε μια τ.μ. X:

1. Κατά πιθανότητα (in probability) εάν, για κάθε $\epsilon > 0$, $\Pr\{|X_n - X| > \epsilon\} \to 0$ για $n \to \infty$.
2. Κατά μέση τετραγωνική τιμή (mean square) εάν $\mathbb{E}(X_n - X)^2 \to 0$.
3. Με πιθανότητα 1 (ή σχεδόν βέβαια) εάν $\Pr\{\lim_{n\to\infty} X_n = X\} = 1$.

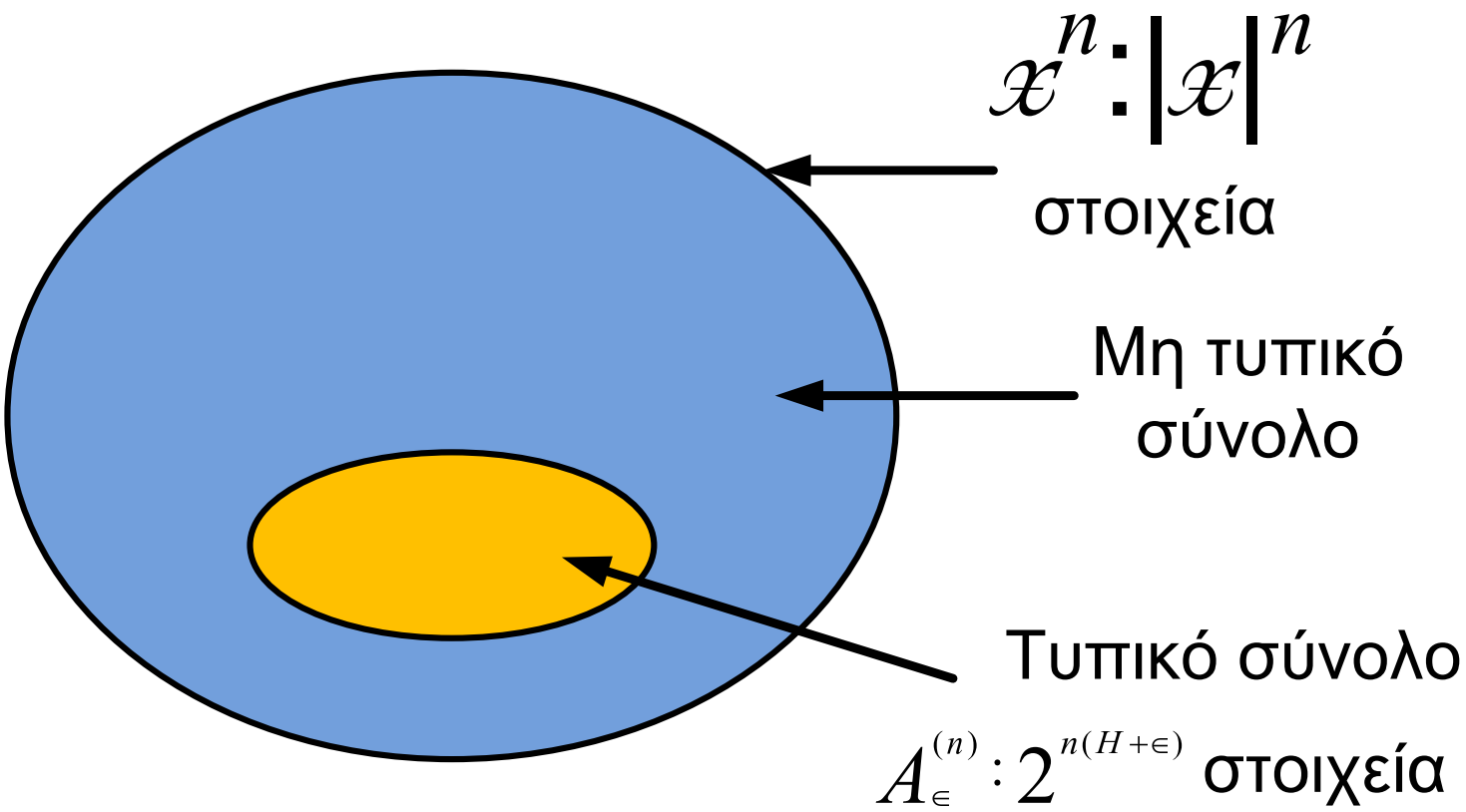
Τυπικό Σύνολο (Typical Set) και ιδιότητες

- Το (ασθενώς) τυπικό σύνολο $A_\epsilon^{(n)}$ 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 (συνέχεια)

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 $n \geq n_0$,

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