ΕΕ728
Προχωρημένα Θέματα Θεωρίας Πληροφορίας
3η διάλεξη
(3η έκδοση, 11/3/2013)

Δημήτρης-Αλέξανδρος Τουμπακάρης

Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών

26 Φεβρουαρίου 2013

Περιεχόμενα 3ης διάλεξης

Αντιστοιχία 3ης διάλεξης με βιβλία Cover & Thomas και El Gamal & Kim

- Βιβλίο Cover & Thomas (2η έκδοση): Κεφ. 2.6 – 2.8, 2.10
- Βιβλίο El Gamal & Kim: Κεφ. 2.1, 2.3

Κυρτές (convex) και κοίλες (concave) συναρτήσεις

- **Ορισμός 3.1.** Μια συνάρτηση $f(x)$ είναι <u>κυρτή</u> ($\cup$) σε διάστημα (a, b) εάν, για κάθε $x_1, x_2 \in (a, b)$ και $0 \leq \lambda \leq 1$,

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

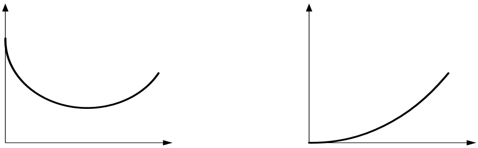
  - Για να είναι μια συνάρτηση κυρτή πρέπει, επίσης, το πεδίο ορισμού της να είναι κυρτό σύνολο. Στον παραπάνω ορισμό έχουμε βασιστεί εμμέσως το γεγονός ότι το διάστημα (a, b) είναι κυρτό: $\forall x_1, x_2 \in (a, b)$, $\lambda x_1 + (1-\lambda)x_2 \in (a, b)$ για $0 \leq \lambda \leq 1$.

- **Ορισμός 3.2.** Μια συνάρτηση $f(x)$ είναι <u>αυστηρώς κυρτή</u> (strictly convex) εάν η ισότητα στην παραπάνω σχέση ισχύει μόνο για $\lambda = 0$ ή $\lambda = 1$ (και παντού αλλού έχουμε $<$).
  - 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 $x \geq 0$).

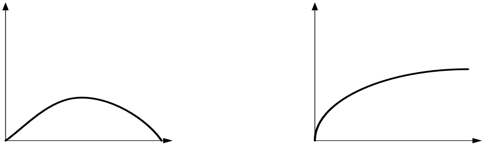
Κυρτές (convex) και κοίλες (concave) συναρτήσεις (συνέχεια)

- Ορισμός 3.3. 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

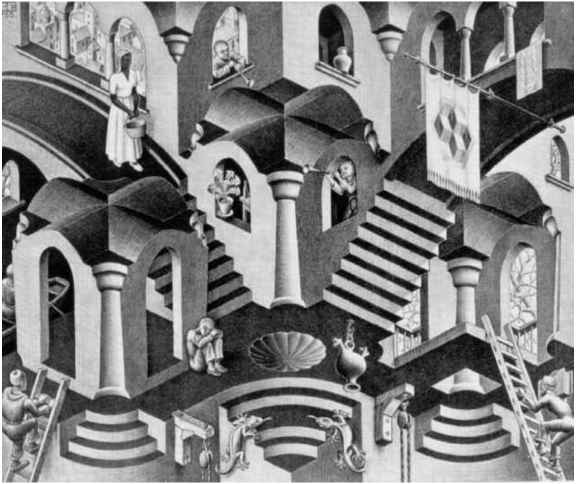


(α) Κυρτές συναρτήσεις

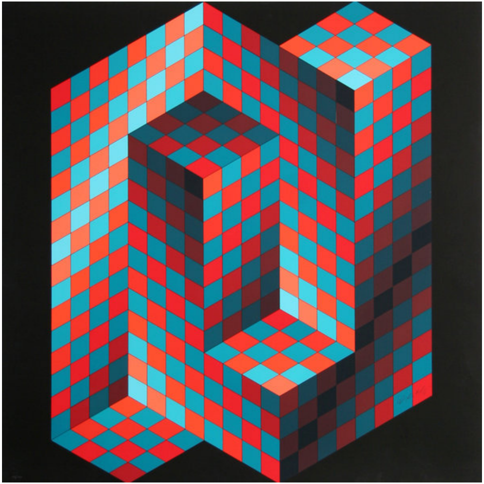


(β) Κοίλες συναρτήσεις

M. C. Escher, Convex and Concave, 1955



V. Vasarely, Gestalt 4, 1970



Ανισότητα Jensen

- **Θεώρημα 3.4.** 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 3.5. (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 $i = 1, 2, \ldots, k-1$.

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

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 3.6.** $H(X) \leq \log|\mathcal{X}|$, 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 (∪) ($(t \log t)'' = \frac{1}{t} \log e > 0$ 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 3.7. 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 3.8. 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 (∪) 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 3.1

- Υποθέτουμε ότι $p(x,y,z) = p(x)p(z)p(y|x,z)$
- Θα αποδείξουμε ότι $I(X;Y|Z) \geq I(X;Y)$.
- 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 3.1 (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 3.1 (3)

- Επομένως, από την Ανισότητα Jensen,

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

Η Ανισότητα Επεξεργασίας Δεδομένων και η Ανισότητα Fano

Ανισότητα Επεξεργασίας Δεδομένων

- Οι X, Y, Z σχηματίζουν *αλυσίδα Markov* ($X \to Y \to Z$) εάν $p(x,y,z) = p(x)p(y|x)p(z|y)$.
  - Ισοδύναμα, $X \to Y \to Z$ εάν και μόνο εάν $p(x,z|y) = p(x|y)p(z|y)$ (δηλαδή, οι x και z είναι υπό συνθήκη ανεξάρτητες *δεδομένης* της y).
  - Ισχύει πάντοτε ότι $X \to Y \to g(Y)$.
- Ανισότητα Επεξεργασίας Δεδομένων (Data Processing Inequality):

Ανισότητα Επεξεργασίας Δεδομένων

Εάν $X \to Y \to Z$, τότε $I(X;Y) \geq I(X;Z)$.

Ανισότητα Επεξεργασίας Δεδομένων (απόδειξη)

- Απόδειξη: Από τον κανόνα αλυσίδας για την αμοιβαία πληροφορία,

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

  λόγω της υπό συνθήκη ανεξαρτησίας των X και Z δεδομένης της Y. Λαμβάνοντας, επίσης, υπόψη ότι $I(X;Y|Z) \geq 0$, προκύπτει η ανισότητα.
- Με τον ίδιο τρόπο μπορούμε, επίσης, να δείξουμε ότι $I(X;Y|Z) \leq I(X;Y)$
- $I(X;Y) \geq I(X;g(Y))$. Συνεπώς, η πληροφορία για τη X που περιέχεται στην Y δεν μπορεί να αυξηθεί με επεξεργασία της Y (αντίθετα, μάλιστα, ενδέχεται να μειωθεί). Ωστόσο, κατάλληλη επεξεργασία της Y ενδέχεται να διευκολύνει την εξαγωγή της πληροφορίας.

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

- Με χρήση της ανισότητας επεξεργασίας δεδομένων.
- Έστω κανάλι με είσοδο X, πίνακα μετάβασης $p(y|x)$ και εξόδους Y.
- Έστω αυθαίρετες κατανομές p_1 και p_2 και I_1 και I_2 η αμοιβαία πληροφορία μεταξύ των X και Y όταν η κατανομή εισόδου είναι η p_1 και p_2, αντίστοιχα. Έστω τυχαία παράμετρος θ, με $0 < \theta < 1$, $p = \theta p_1 + (1-\theta)p_2$ και I η αντίστοιχη αμοιβαία πληροφορία. Θα δείξουμε ότι
$$\theta I_1 + (1-\theta)I_2 \leq I.$$

Η $I(X;Y)$ είναι κοίλη ($\cap$) συνάρτηση της $p(x)$ για δεδομένη $p(y|x)$ – Εναλλακτική Απόδειξη (2)

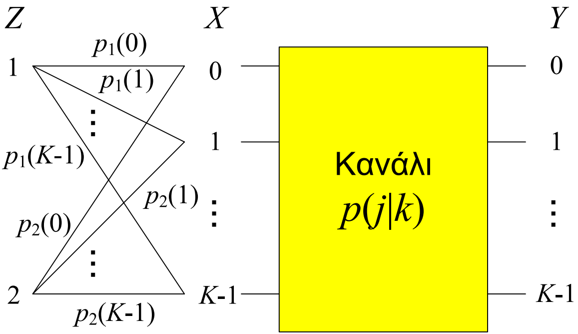
- Μπορούμε να υποθέσουμε ότι οι p_1 και p_2 είναι υπό συνθήκη κατανομές που εξαρτώνται από μια δυαδική τ.μ. Z:

$$p_1(x) = p_{X|Z}(x|1), \qquad p_2(x) = p_{X|Z}(x|2)$$

- Θέτουμε $p_Z(1) = \theta$ και $p_Z(2) = 1 - \theta$.

Η $I(X;Y)$ είναι κοίλη ($\cap$) συνάρτηση της $p(x)$ για δεδομένη $p(y|x)$ – Εναλλακτική Απόδειξη (3)

Το πρόβλημα φαίνεται στο παρακάτω σχήμα.



Παρατηρούμε ότι $Z \to X \to Y$ και $p(y|x,z) = p(y|x)$.
Επίσης, $\theta I_1 + (1-\theta)I_2 = I(X;Y|Z)$ και $I = I(X;Y)$.

Η $I(X;Y)$ είναι κοίλη ($\cap$) συνάρτηση της $p(x)$ για δεδομένη $p(y|x)$ – Εναλλακτική Απόδειξη (4)

- Δεδομένου ότι οι Z και Y είναι υπό συνθήκη ανεξάρτητες δεδομένης της X, $I(Y;Z|X) = 0$.
- Επίσης, όπως και στην απόδειξη της ανισότητας επεξεργασίας δεδομένων,

$$I(Y;X,Z) = I(Y;Z) + I(Y;X|Z) = I(Y;X) + I(Y;Z|X) \Rightarrow$$
$$I(Y;Z) + I(Y;X|Z) = I(Y;X) \Rightarrow$$
$$I(Y;X|Z) = I(X;Y|Z) \leq I(Y;X).$$

- Με παρόμοιο τρόπο μπορεί να αποδειχτεί ότι η $I(X;Y)$ είναι κυρτή ($\cup$) συνάρτηση της $p(y|x)$ για δεδομένη $p(x)$ (Gallager Theorem 4.4.3).

Η Ανισότητα Fano

Εκτίμηση τιμής τυχαίας μεταβλητής

- Σκοπός της επικοινωνίας είναι ο δέκτης να λάβει την πληροφορία που του στέλνει ο πομπός μέσω ενός καναλιού.
- Έστω ότι η τ.μ. Y περιέχει κάποια πληροφορία για τη X (οπότε οι X και Y δεν είναι ανεξάρτητες και $I(X;Y) > 0$).
- Εκτιμητής (estimator): Μια συνάρτηση της Y η οποία παράγει μια εκτίμηση (estimate) για τη X: $\hat{X} = g(Y)$.
- Ο εκτιμητής μπορεί να είναι ντετερμινιστικός (deterministic) ή στοχαστικός.
- Θέλουμε να βρούμε ποια είναι η πιθανότητα η εκτίμηση $\hat{X}$ να μην ισούται με την πραγματική τιμή της τ.μ. X που μετέδωσε ο πομπός.
- Ορίζουμε την Πιθανότητα Σφάλματος $P_e \triangleq \Pr\{\hat{X} \neq X\}$.

Εκτίμηση τιμής τυχαίας μεταβλητής (συνέχεια)

- Προφανώς, εάν $H(X|Y) = 0$, υπάρχει εκτιμητής ο οποίος παράγει εκτιμήσεις με $P_e = 0$.
- Διαισθητικά περιμένουμε ότι μικρές τιμές της $H(X|Y)$ θα οδηγούν σε εκτιμήσεις με μικρή P_e (εφόσον, βέβαια, χρησιμοποιηθεί καλός εκτιμητής).
- Η ανισότητα Fano δίνει ένα *κάτω φράγμα* για την P_e συναρτήσει της $H(X|Y)$.

Ανισότητα Fano

- Για κάθε εκτιμητή τέτοιο ώστε $X \to Y \to \hat{X}$,

Ανισότητα Fano

$$H(P_e) + P_e \log |\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y),$$

όπου $H(P_e) = -P_e \log P_e - (1 - P_e) \log(1 - P_e)$.

- Παρατηρήστε ότι ο εκτιμητής δεν είναι, κατ' ανάγκη, ντετερμινιστική συνάρτηση της Y. Επίσης, $P_e = 0 \Rightarrow H(X|Y) = 0$.

Ανισότητα Fano (συνέχεια)

- Θέτοντας $H(P_e) = \max_p H(p) = 1$ προκύπτει το λιγότερο ακριβές κάτω φράγμα,

$$1 + P_e \log|\mathcal{X}| \geq H(X|Y) \Rightarrow P_e \geq \frac{H(X|Y) - 1}{\log|\mathcal{X}|}.$$

- Θα χρησιμοποιήσουμε την ανισότητα Fano στην απόδειξη του Θεωρήματος Κωδικοποίησης Καναλιού (αντίστροφο).

Απόδειξη Ανισότητας Fano

(Cover & Thomas Theorem 2.10.1)

- Έστω η τ.μ. E που υποδηλώνει εάν έχει εμφανιστεί σφάλμα ή όχι στην εκτίμηση της X

$$E = \begin{cases} 1 & \text{εάν} \quad \hat{X} \neq X, \\ 0 & \text{εάν} \quad \hat{X} = X. \end{cases}$$

- Αναπτύσσουμε την $H(E, X|\hat{X})$ με χρήση του κανόνα αλυσίδας για την εντροπία:

$$\begin{aligned} H(E, X|\hat{X}) &= H(X|\hat{X}) + \underbrace{H(E|X, \hat{X})}_{=0} \\ &= \underbrace{H(E|\hat{X})}_{\leq H(E) = H(P_e)} + \underbrace{H(X|E, \hat{X})}_{\leq P_e \log |\mathcal{X}|}. \end{aligned}$$

  - $H(E|X, \hat{X}) = 0$ γιατί εάν ξέρουμε τις τιμές των $\hat{X}$ και X γνωρίζουμε εάν έχει εμφανιστεί σφάλμα εκτίμησης.

Απόδειξη Ανισότητας Fano (2)

$$H(E,X|\hat{X}) = H(X|\hat{X}) + \underbrace{H(E|X,\hat{X})}_{=0}$$
$$= \underbrace{H(E|\hat{X})}_{\leq H(E)=H(P_e)} + \underbrace{H(X|E,\hat{X})}_{\leq P_e \log|\mathcal{X}|}.$$

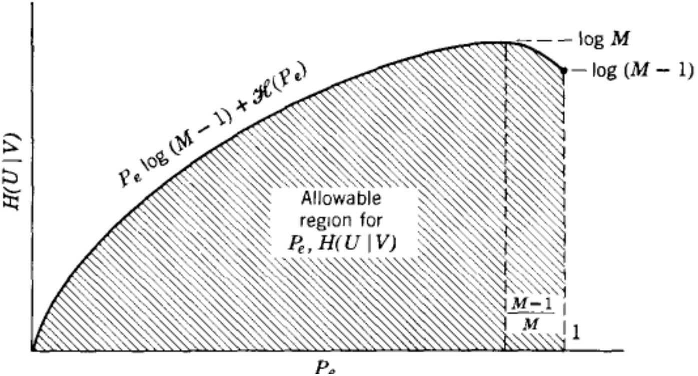
- $H(E|\hat{X}) \leq H(E)$. Δεδομένου ότι η πιθανότητα σφάλματος ($E=1$) ισούται με P_e, η τ.μ. ακολουθεί κατανομή Bernoulli με παράμετρο P_e και $H(E)=H(P_e)$.
- $H(X|E,\hat{X}) = \Pr(E=0)H(X|\hat{X},E=0)+\Pr(E=1)H(X|\hat{X},E=1) \leq (1-P_e)0+P_e\log|\mathcal{X}|$, δεδομένου ότι εάν δεν υπάρχει σφάλμα εκτίμησης $X=\hat{X}$, ενώ η χειρότερη περίπτωση εάν έχει συμβεί σφάλμα είναι η X να ακολουθεί ομοιόμορφη κατανομή.

- Επομένως, $H(P_e)+P_e\log|\mathcal{X}| \geq H(X|\hat{X})$.

Απόδειξη Ανισότητας Fano (3)

- $H(P_e) + P_e \log|\mathcal{X}| \geq H(X|\hat{X})$.
- Δεδομένου ότι $X \to Y \to \hat{X}$,
  $I(X;Y) \geq I(X;\hat{X}) \Rightarrow H(X) - H(X|Y) \geq H(X) - H(X|\hat{X}) \Rightarrow$
  $H(X|\hat{X}) \geq H(X|Y)$.
  Συνεπώς,
  $$H(P_e) + P_e \log|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y).$$
- Εάν απαιτήσουμε η εκτιμώμενη τιμή $\hat{X}$ να ανήκει στο σύνολο $\mathcal{X}$, $H(X|E,\hat{X}) \leq P_e \log(|\mathcal{X}| - 1)$ και
  $$H(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|\hat{X}) \geq H(X|Y).$$

Επιτρεπτή περιοχή για $P_e, H(X|Y)$



c R. G. Gallager, *Information Theory and Reliable Communication*, 1968