EE728
Προχωρημένα Θέματα Θεωρίας Πληροφορίας
2η και 3η διάλεξη

Δημήτρης-Αλέξανδρος Τουμπακάρης

Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών

4 και 11 Μαρτίου 2014

Περιεχόμενα 2ης και 3η2 διάλεξης

Αντιστοιχία 2ης και 3ης διάλεξης με βιβλία Cover & Thomas και El Gamal & Kim

- Βιβλίο Cover & Thomas (2η έκδοση): Κεφ. 3, Κεφ. 2.1 – 2.8, 2.10
- Βιβλίο El Gamal & Kim: Κεφ. 2.1, 2.3, 2.4, 3.5

Τυπικό Σύνολο (Typical Set)

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

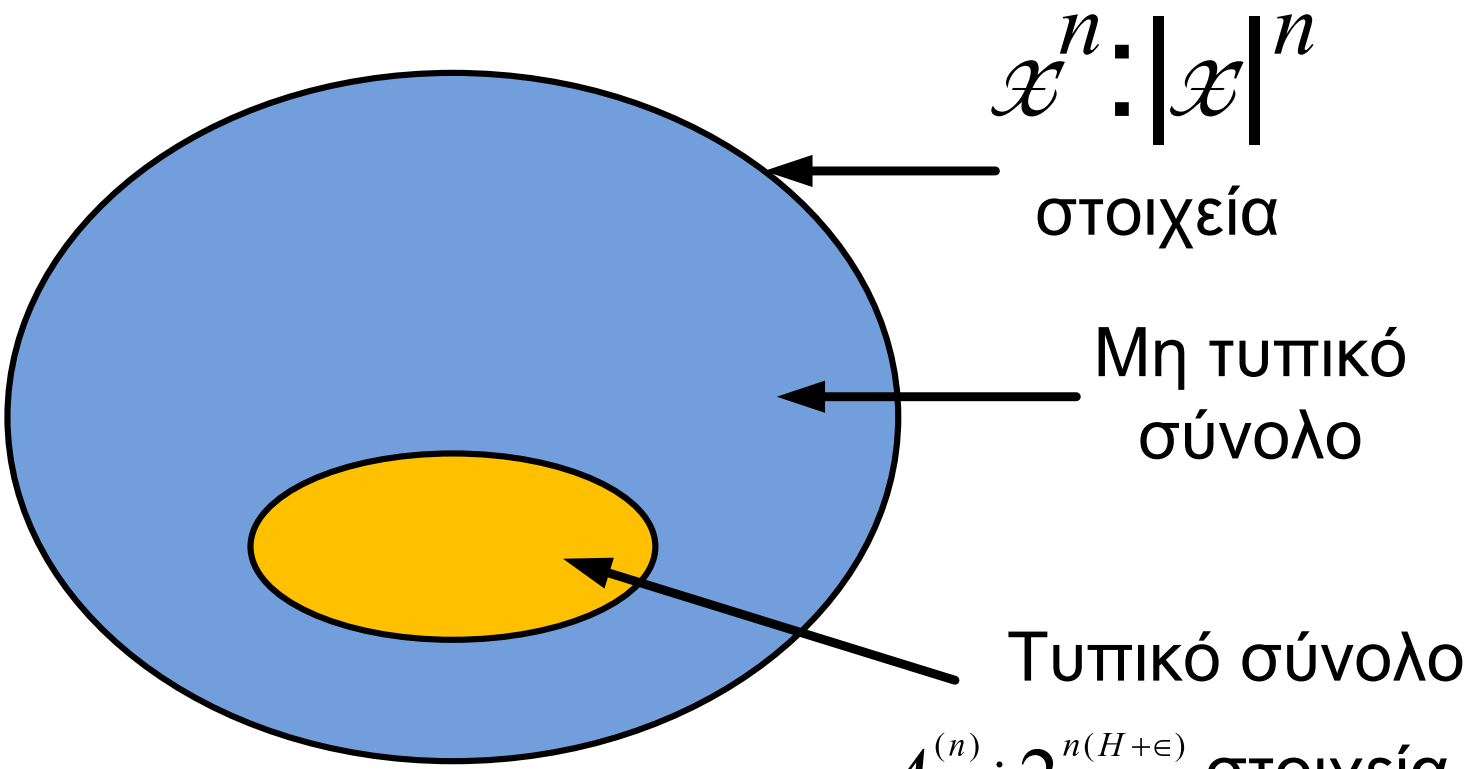
$$2^{-n(H(X)+\epsilon)} \leq p(x_1, x_2, \ldots, x_n) \leq 2^{-n(H(X)-\epsilon)}.$$

- Θεωρούμε ότι τα στοιχεία, x_i, των ακολουθιών είναι i.i.d. και ακολουθούν κατανομή $p(x)$.

Ιδιότητες Τυπικού Συνόλου

- 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 (πληθικότητα – cardinality) 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 ασθενώς 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.

Αποδείξεις ιδιοτήτων Τυπικού Συνόλου (3)

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 $n \geq n_0$,

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

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

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

Παράδειγμα 2.1 (Cover & Thomas Problem 3.6)

- Έστω οι ανεξάρτητες και ομοίως κατανεμημένες (i.i.d.) 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 (θεωρούμε μικρό ϵ). 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.

$$\text{Ισχυρή Τυπικότητα} \Rightarrow \text{Ασθενής Τυπικότητα}$$

- Το αντίστροφο δεν ισχύει, όπως είδαμε στο παράδειγμα πηγής $\text{Bern}(1/2)$ χωρίς μνήμη.
- Η ισχυρή τυπικότητα είναι πιο ευέλικτη από την ασθενή. Επίσης, για την απόδειξη κάποιων θεωρημάτων της Θεωρίας Πληροφορίας δεν επαρκεί η Ασθενής Τυπικότητα και απαιτείται χρήση Ισχυρής Τυπικότητας.
- Μπορούμε να αποδείξουμε τις ίδιες ιδιότητες για τις ισχυρώς τυπικές ακολουθίες όπως και για τις ασθενώς τυπικές με παρόμοιο τρόπο.

Σθεναρή Τυπικότητα

- Συχνά, το 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.

- Το είδος αυτό τυπικότητας ονομάζεται *σθεναρή* (robust typicality).
- Παρατηρήστε ότι

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

- Στη συνέχεια, όταν αναφερόμαστε σε ισχυρή τυπικότητα θα εννοούμε τη σθεναρή τυπικότητα.

Λήμμα Τυπικού Μέσου (Typical Average Lemma)

- Όπως προαναφέρθηκε, το AEP αποτελεί ειδική περίπτωση του Λήμματος Τυπικού Μέσου.

Typical Average Lemma

Έστω ότι η ακολουθία $x_1^n \in \mathcal{T}_\epsilon^{(n)}$. Για οποιαδήποτε *μη αρνητική* συνάρτηση $g(x)$ με πεδίο ορισμού το $\mathcal{X}$,

$$(1-\epsilon)\mathbb{E}\left[g(X)\right] \leq \frac{1}{n}\sum_{i=1}^{n} g(x_i) \leq (1+\epsilon)\mathbb{E}\left[g(X)\right].$$

- Απόδειξη: Προκύπτει εύκολα από τον ορισμό της σθεναρής τυπικότητας (κάντε το ως άσκηση).

Ιδιότητες σθεναρώς τυπικού συνόλου

- Για το σθεναρώς τυπικό συνολο ισχύουν οι ίδιες ιδιότητες με το ασθενώς τυπικό, με τη διαφορά ότι στα παρακάτω $\delta(\epsilon) = \epsilon H(X)$.
- Δηλαδή,
  1. Εάν $(x_1, x_2, \ldots, x_n) \in \mathcal{T}_\epsilon^{(n)}$, $H(X) - \delta(\epsilon) \leq -\frac{1}{n} \log p(x_1, x_2, \ldots, x_n) \leq H(X) + \delta(\epsilon)$.
  2. $\Pr\left\{\mathcal{T}_\epsilon^{(n)}\right\} > 1 - \epsilon$ για n μεγαλύτερο από κάποια τιμή n_0.
  3. $\left|\mathcal{T}_\epsilon^{(n)}\right| \leq 2^{n(H(X)+\delta(\epsilon))}$,
  4. $\left|\mathcal{T}_\epsilon^{(n)}\right| \geq (1-\epsilon)2^{n(H(X)-\delta(\epsilon))}$, για n μεγαλύτερο από κάποια τιμή n_0.

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 ($\log \to \log_2$) ή 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: $p(x) = \frac{1}{|\mathcal{X}|}$ για όλα τα $x \in \mathcal{X}$. Αποδείχτηκε 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):

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

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

- Μερικές φορές ο ρυθμός εντροπίας συμβολίζεται με $\bar{H}(X)$.

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.
- $D(p||q) \geq 0$. Αποδείχτηκε 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 $I(X;Y) = \mathbb{E}_Y[I(X;Y=y)] \geq 0$.

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)

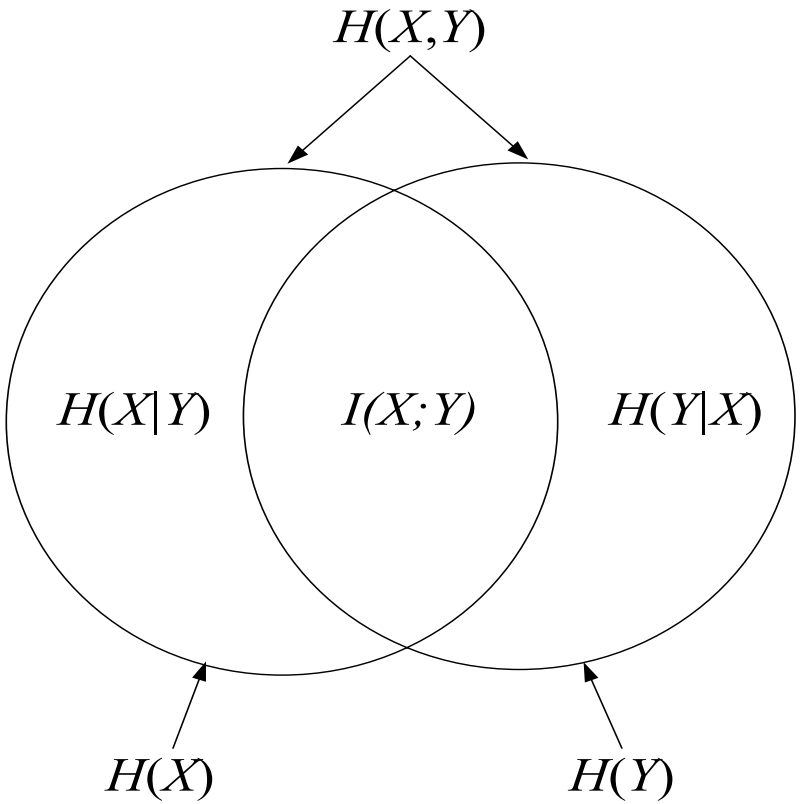
- $I(X;X) = H(X) - H(X|X) = H(X)$. Η X περιέχει όλη την πληροφορία για τον εαυτό της.
- Κανόνας αλυσίδας για την αμοιβαία πληροφορία:

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

- 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.



Κυρτές (convex) και κοίλες (concave) συναρτήσεις

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

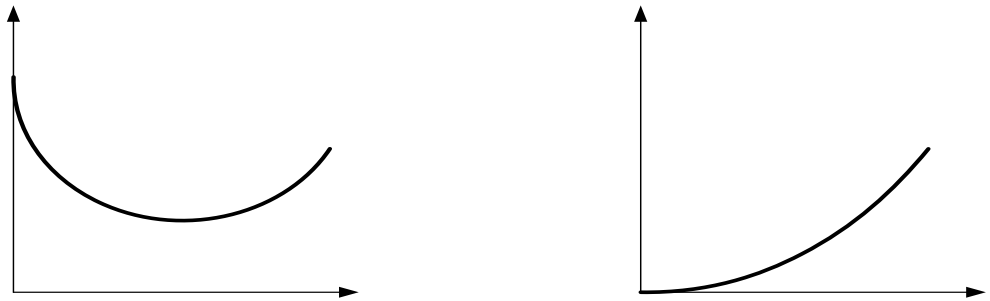
$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

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

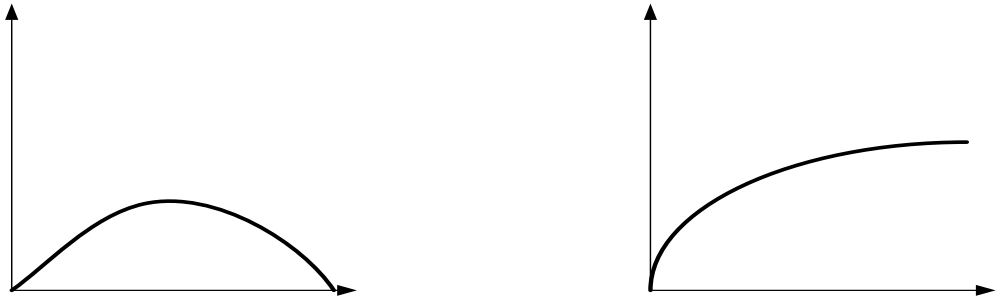
Κυρτές (convex) και κοίλες (concave) συναρτήσεις (συνέχεια)

- Ορισμός 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

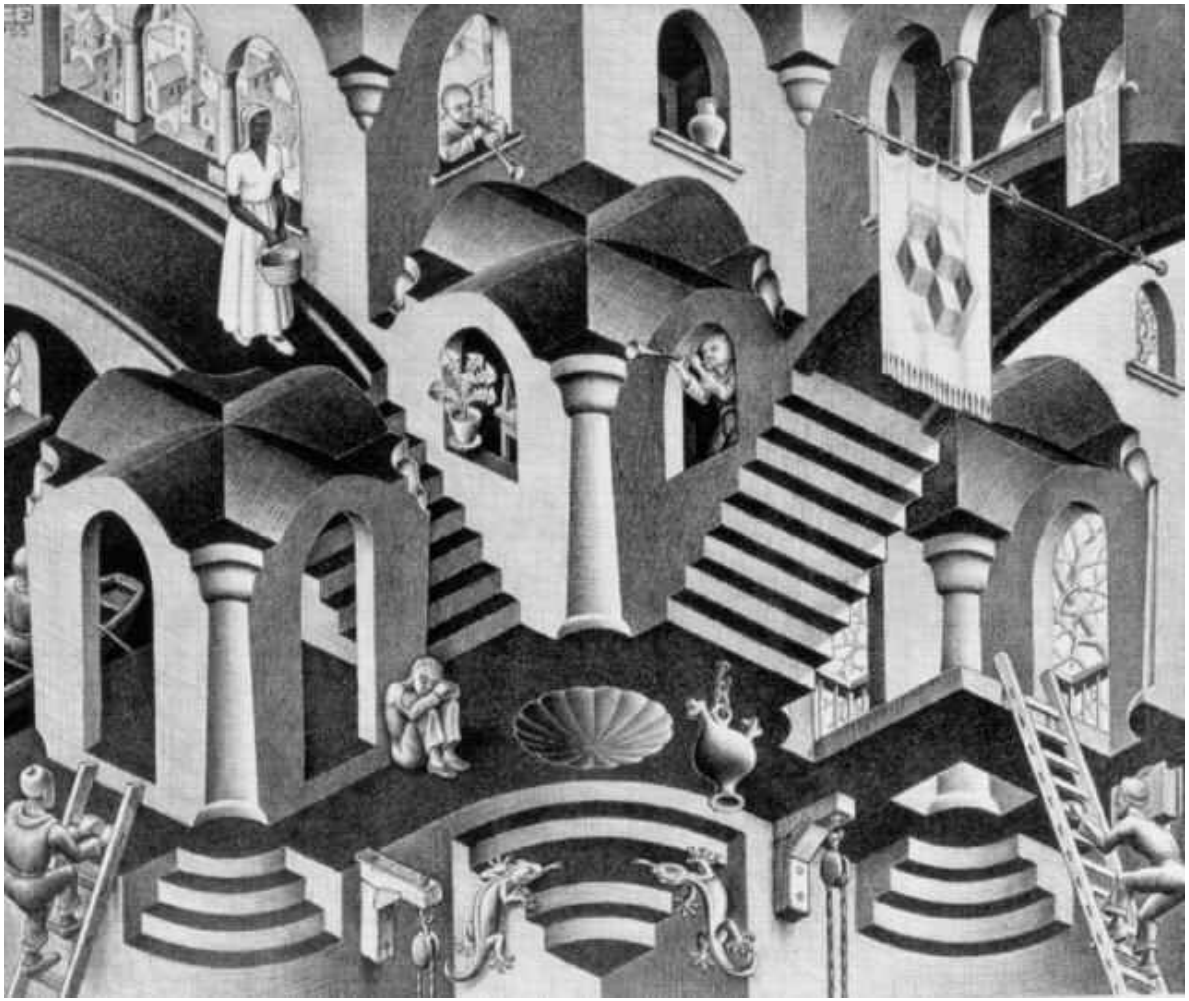


(α) Κυρτές συναρτήσεις

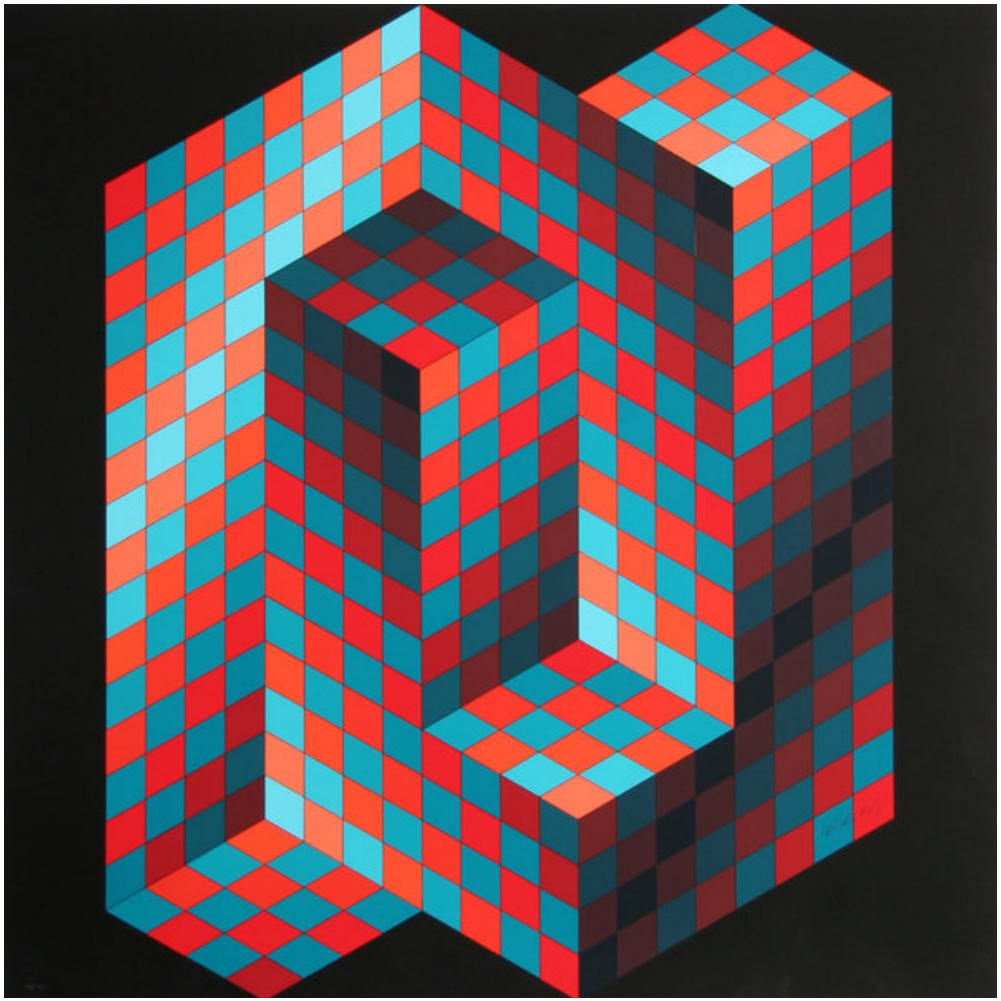


(β) Κοίλες συναρτήσεις

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 ($\cup$) 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 $I(X;Y|Z) \geq I(X;Y)$.
- 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 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,

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

Παράδειγμα 2.3 (4)

- Εναλλακτικά, και πολύ πιο απλά,

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) \\ &\overset{(a)}{=} H(X) - H(X|Y,Z) \\ &\overset{(b)}{\geq} H(X) - H(X|Y) = I(X;Y) \end{aligned}$$

(a) από την ανεξαρτησία των X και Z, (b) conditioning reduces the entropy.

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)

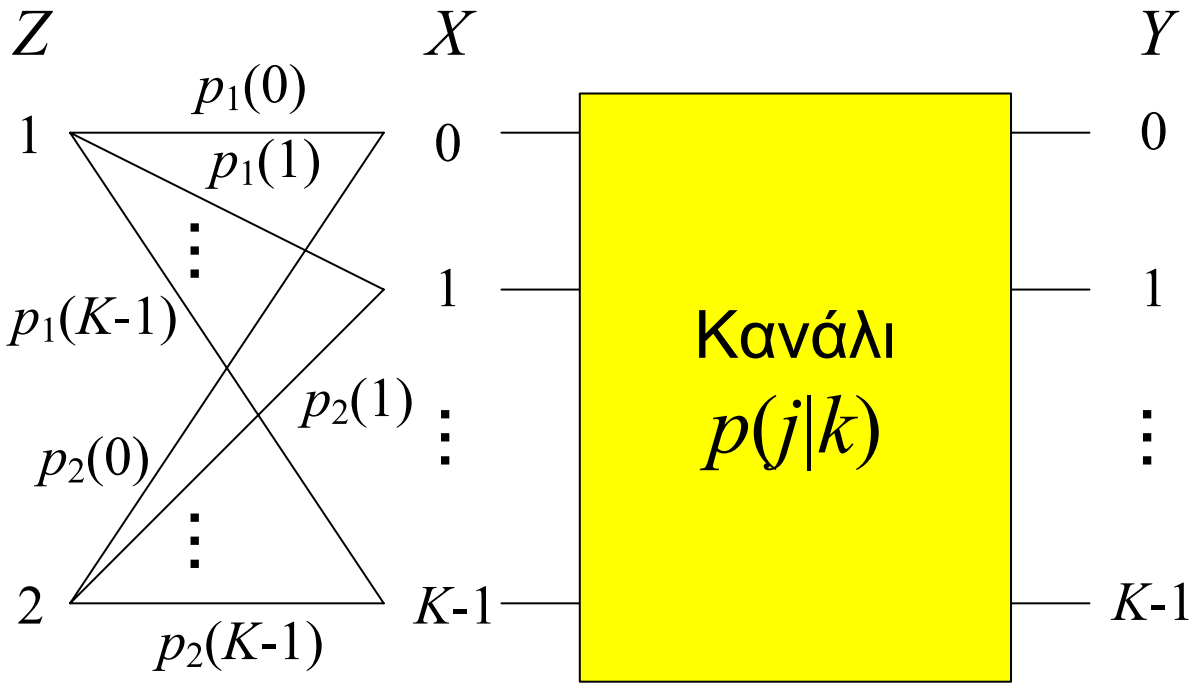
- Μπορούμε να υποθέσουμε ότι οι p_1 και p_2 είναι υπό συνθήκη κατανομές που εξαρτώνται από μια δυαδική τ.μ. Z:

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

- 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 και στην απόδειξη του Θεωρήματος Κωδικοποίησης Πηγής (αντίστροφο).
- Γενικώς, η ανισότητα Fano χρησιμοποιείται στις περισσότερες αποδείξεις αντιστρόφων (converse proofs).

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)

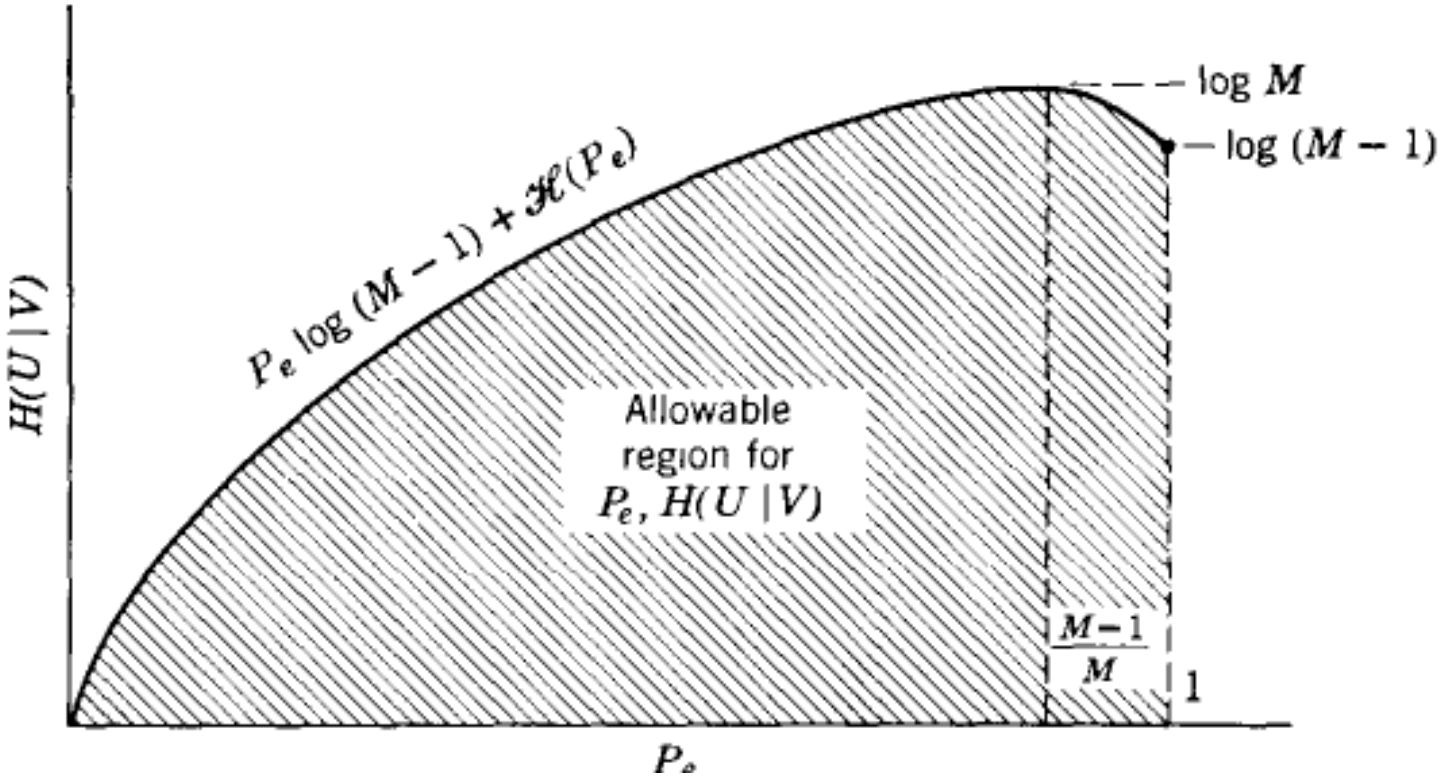
- $H(P_e) + P_e \log |\mathcal{X}| \geq H(X|\hat{X})$.
- Δεδομένου ότι $X \to Y \to \hat{X}$,
  $I(X;Y) \geq I(X;\hat{X}) \Rightarrow H(X) - H(X|Y) \geq H(X) - H(X|\hat{X}) \Rightarrow$
  $H(X|\hat{X}) \geq H(X|Y)$.
  Συνεπώς,

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

- 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 $P_e, H(X|Y)$



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

Κωδικοποίηση Σταθερού Μήκους και Θεώρημα Κωδικοποίησης Πηγής

Κωδικοποίηση Σταθερού Μήκους

- Έστω, ανεξάρτητες, ομοίως κατανεμημένες (i.i.d) τ.μ. $X_i \sim p(x)$. Θέλουμε να βρούμε αποδοτική περιγραφή ακολουθιών $X_1, X_2, \ldots, X_n$ των τ.μ.
- Χωρίζουμε όλες τις $|\mathcal{X}|^n$ πιθανές ακολουθίες σε 2 σύνολα: Το τυπικό σύνολο $A_\epsilon^{(n)}$ και το μη τυπικό σύνολο $A_\epsilon^{(n)^c} = \mathcal{X}^n - A_\epsilon^{(n)}$.
- Κατασκευή βιβλίου κωδίκων (codebook): Διατάσσουμε όλες τις τυπικές ακολουθίες (π.χ. με αλφαβητική σειρά) και σε κάθε ακολουθία αντιστοιχίζουμε μία κωδική λέξη μήκους L.
- Δεδομένου ότι το τυπικό σύνολο περιέχει το πολύ $2^{n(H+\epsilon)}$ ακολουθίες (σύμφωνα με την Ιδιότητα 3), χρειαζόμαστε το πολύ $L = n(H+\epsilon) + 1$ bits για να τις αναπαραστήσουμε (το επιπλέον 1 bit οφείλεται στο ότι ενδέχεται η ποσότητα $n(H+\epsilon)$ να μην είναι ακέραιος).
- Όλες οι κωδικές λέξεις έχουν το *ίδιο* μήκος L.

Κωδικοποίηση Σταθερού Μήκους (2)

- Σχηματίζουμε ακολουθία μήκους $n > n_0$ από τα σύμβολα X_i της πηγής που θέλουμε να κωδικοποιήσουμε.
- Κωδικοποίηση (encoding):
  - Εάν η ακολουθία είναι τυπική, την κωδικοποιούμε με την κωδική λέξη μήκους L του βιβλίου κωδίκων.
  - Εάν η ακολουθία δεν είναι τυπική, η κωδικοποίηση αποτυγχάνει.
  - Μπορούμε να ελαπτώσουμε την πιθανότητα αποτυχίας, ϵ, όσο θέλουμε αυξάνοντας το μήκος, n, των ακολουθιών που κωδικοποιούμε.
- Επομένως, μπορούμε να κωδικοποιήσουμε με χρήση $L/n = (H + \epsilon) + 1/n$ bits/σύμβολο πηγής και να διασφαλίσουμε ότι η πιθανότητα αποτυχίας είναι μικρότερη του ϵ.

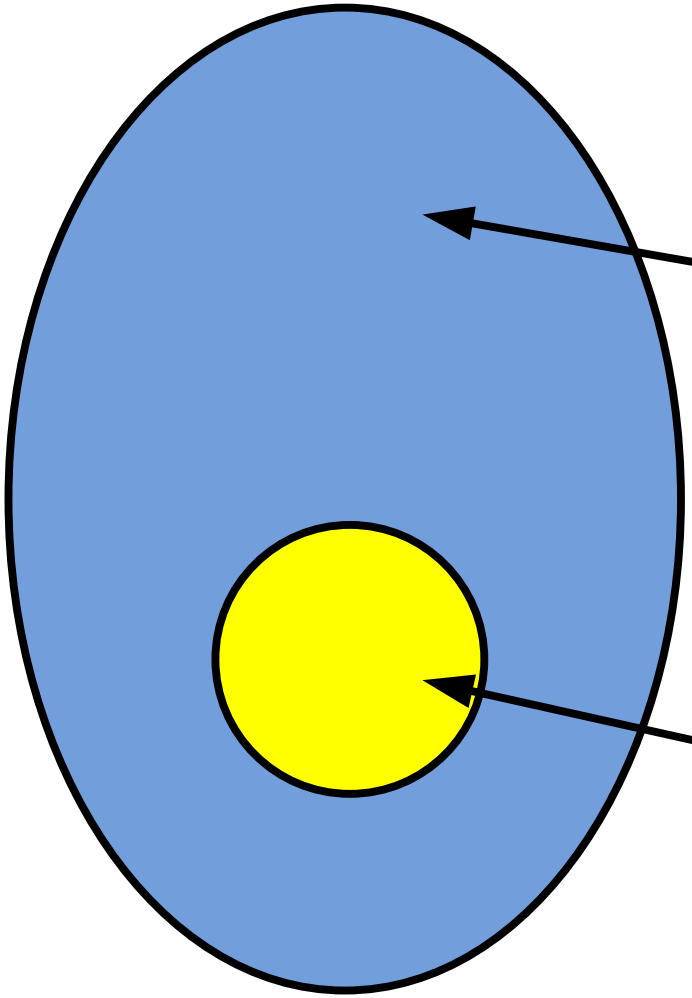
Κωδικοποίηση Σταθερού Μήκους (3)

- Με μία μικρή αλλαγή στον τρόπο κωδικοποίησης μπορούμε να διασφαλίσουμε ότι η πιθανότητα αποτυχίας κωδικοποίησης είναι ακριβώς ίση με 0.
- Ωστόσο, στην περίπτωση αυτή, η κωδικοποίηση δεν είναι σταθερού μήκους.
- Παρατηρούμε ότι για να περιγράψουμε τις ακολουθίες του μη τυπικού συνόλου χρειαζόμαστε το πολύ $n \log \left|\mathcal{X}\right| + 1$ bits.
- Διατηρούμε το βιβλίο κωδίκων των τυπικών ακολουθιών και προσθέτουμε και ένα βιβλίο κωδίκων για τις μη τυπικές ακολουθίες.
- Το βιβλίο κωδίκων για τις μη τυπικές ακολουθίες μπορεί να είναι τετριμμένο, δηλαδή να μη συμπιέζουμε την ακολουθία.

Κωδικοποίηση Σταθερού Μήκους (4)

- Κωδικοποίηση:
  - Σχηματίζουμε ακολουθία μήκους $n > n_0$ από τα σύμβολα X_i της πηγής που θέλουμε να κωδικοποιήσουμε.
  - Εάν η ακολουθία είναι τυπική, χρησιμοποιούμε πρόθεμα 0 και το βιβλίο κωδίκων των τυπικών ακολουθιών (μήκους L). Επομένως, χρειαζόμαστε $L + 1 = n(H(X) + \epsilon) + 2$ bits.
  - Αλλιώς, αν η ακολουθία είναι μη τυπική, χρησιμοποιούμε πρόθεμα 1 και, στη, συνέχεια, την ίδια την ακολουθία (χωρίς να τη συμπιέσουμε). Επομένως, χρειαζόμαστε $n \log |\mathcal{X}| + 2$ bits.

Κωδικοποίηση Σταθερού Μήκους με χρήση τυπικού συνόλου



Μη τυπικό σύνολο

Περιγραφή: $n\log|\mathcal{X}|+2$ bits

Τυπικό σύνολο

Περιγραφή: $n(H+\varepsilon)+2$ bits

Κωδικοποίηση Σταθερού Μήκους (συνέχεια)

- Το μέσο μήκος της κωδικής λέξης ισούται με

$$\begin{aligned}\mathbb{E}[l(X^n)] &= \sum_{x^n} p(x^n) l(x^n) = \sum_{x^n \in A_\epsilon^{(n)}} p(x^n) l(x^n) + \sum_{x^n \in A_\epsilon^{(n)c}} p(x^n) l(x^n) \\ &\leq \sum_{x^n \in A_\epsilon^{(n)}} p(x^n) \left((nH + \epsilon) + 2\right) + \sum_{x^n \in A_\epsilon^{(n)c}} p(x^n)(n \log |\mathcal{X}| + 2) \\ &= \Pr\left\{A_\epsilon^{(n)}\right\} \left[(nH + \epsilon) + 2\right] + \Pr\left\{A_\epsilon^{(n)^c}\right\} \left[n \log |\mathcal{X}| + 2\right] \\ &\leq (nH + \epsilon) + 2 + \epsilon(n \log |\mathcal{X}| + 2) = n(H + \epsilon').\end{aligned}$$

- Το $\epsilon' = \epsilon + \epsilon \log |\mathcal{X}| + \frac{2+\epsilon}{n}$ μπορεί να γίνει αυθαίρετα μικρό επιλέγοντας κατάλληλη τιμή του n και του ϵ (το οποίο εξαρτάται από το n).
- Συνεπώς, $\mathbb{E}\left[\frac{1}{n} l(X^n)\right] \leq H(X) + \epsilon'$ για $n > n_1$.

Παρατηρήσεις

- Δείξαμε ότι υπάρχει (τουλάχιστον ένας) τρόπος να συμπιέσουμε μια ακολουθία μήκους n με χρήση $\sim nH$ bits (αντί για $n \log |\mathcal{X}|$).
- Η σημαντική παρατήρηση είναι ότι, καθώς το μήκος της ακολουθίας τείνει στο άπειρο, η πιθανότητα να εμφανιστεί μη τυπική ακολουθία τείνει στο 0. Μάλιστα, η κωδικοποίηση των μη τυπικών ακολουθιών έγινε χωρίς να ληφθεί πρόνοια να είναι όσο το δυνατόν πιο αποδοτική (χρησιμοποιώντας, π.χ. $n \log \left| A_\epsilon^{(n)^c} \right|$ bits).
- Παρατηρήστε ότι στο όριο το τυπικό σύνολο περιέχει πολύ μικρό ποσοστό των ακολουθιών (το μέγεθός του είναι $\sim 2^{nH}$ οπότε περιέχει ποσοστό $\sim 2^{n(H(X)-\log|\mathcal{X}|)} \to 0$ για $n \to \infty$ εάν $H(X) < \log|\mathcal{X}| - \epsilon$). Ωστόσο, τα στοιχεία του περιέχουν (σχεδόν) όλη την πιθανότητα!

Παρατηρήσεις (συνέχεια)

- Δε χάσαμε καθόλου πληροφορία με την κωδικοποίηση, δεδομένου ότι σε κάθε ακολουθία αντιστοιχίσαμε μια μοναδική κωδική λέξη.
- Ωστόσο, παρατηρούμε ότι, για να συμπιέσουμε αποδοτικά, χρειαζόμαστε μεγάλα μήκη ακολουθιών και, επομένως, δημιουργούνται μεγάλες απαιτήσεις σε καθυστέρηση και μνήμη.
- Θα αποδείξουμε ότι δεν υπάρχει κώδικας χωρίς απώλειες που επιτυγχάνει συμπίεση με λιγότερα bits ανά σύμβολο από την εντροπία (Αντίστροφο Θεωρήματος Κωδικοποίησης Πηγής).

Θεώρημα Κωδικοποίησης Πηγής

Θεώρημα Κωδικοποίησης Πηγής

- Είδαμε ότι, για πηγή χωρίς μνήμη, μπορούμε να πετύχουμε συμπίεση αυθαίρετα κοντά στην εντροπία αυξάνοντας το μήκος των κωδικοποιούμενων ακολουθιών (εκμεταλλευόμενοι το AEP).
- Στο μάθημα "Θεωρία Πληροφορίας" είδαμε, επίσης, ότι, για βέλτιστους κώδικες μεταβλητού μήκους και πηγή χωρίς μνήμη, $H(X) \leq \mathbb{E}[l^*] < H(X)+1 \Rightarrow H(X^n) \leq \mathbb{E}[\tilde{l}^*] < H(X^n)+1 \Rightarrow nH(X) \leq \mathbb{E}[\tilde{l}^*] < nH(X)+1 \Rightarrow H(X) \leq \mathbb{E}[\tilde{l}^*]/n < H(X)+1/n$.
- Επομένως, υπάρχει και δεύτερος τρόπος να συμπιέσουμε κοντά στην εντροπία, αυτή τη φορά με κώδικα μεταβλητού μήκους.

Θεώρημα Κωδικοποίησης Πηγής (2)

- Οι δύο τρόποι κωδικοποίησης που προαναφέρθηκαν αποτελούν αποδείξεις της *επιτευξιμότητας* (achievability) του Θεωρήματος Κωδικοποίησης Πηγής για πηγές χωρίς μνήμη (το οποίο, επίσης, ονομάζεται ευθύ μέρος του Θεωρήματος).
- Ωστόσο, για να αποδειχθεί το Θεώρημα Κωδικοποίησης Πηγής πρέπει, επίσης, να δείξουμε ότι δεν υπάρχει τρόπος να συμπιέσουμε περισσότερο τα σύμβολα της πηγής (αντίστροφο (converse) του Θεωρήματος).
- Ένας άλλος τρόπος να το σκεφτούμε είναι ο εξής: Η επιτευξιμότητα μας δίνει ένα άνω φράγμα για το μήκος της περιγραφής της συμπιεσμένης ακολουθίας. Αν βρούμε ένα κάτω φράγμα το οποίο ταυτίζεται με το άνω φράγμα έχουμε αποδείξει το Θεώρημα.
- Θα αποδείξουμε ότι, εάν προσπαθήσουμε να συμπιέσουμε με μέσο μήκος μικρότερο από την εντροπία, η πιθανότητα αδυναμίας αποκωδικοποίησης $P_e \to 1$.

Θεώρημα Κωδικοποίησης Πηγής (3) – αντίστροφο

- Δείτε π.χ. R. Gallager, *Information Theory and Reliable Communication*, R. W. Yeung, *A First Course in Information Theory*.
- Έστω ότι το μήκος της αρχικής (προς συμπίεση) ακολουθίας ισούται με n. Θεωρούμε δυαδικές ακολουθίες (αν και η απόδειξη γενικεύεται εύκολα). Έστω ότι η ακολουθία συμπιέζεται με χρήση L bits, όπου $L < n[H(X) - \zeta]$, $\zeta > 0$ και ότι το ζ δε μεταβάλλεται με το n. Επομένως, μπορούμε να ανακατασκευάσουμε το πολύ $M = 2^{n(H(X)-\zeta)}$ ακολουθίες στην έξοδο του αποκωδικοποιητή.
- Έστω ότι αντιστοιχίζουμε κάποιες από τις M κωδικές λέξεις σε τυπικές ακολουθίες και κάποιες σε μη τυπικές.
- Το άθροισμα των μαζών πιθανότητας των τυπικών ακολουθιών που μπορούμε να κωδικοποιήσουμε δεν μπορεί να υπερβαίνει την τιμή

$$2^{n[H(X)-\zeta]}2^{-n[H(x)-\epsilon]} = 2^{-n[\zeta-\epsilon]}.$$

Θεώρημα Κωδικοποίησης Πηγής (4) – αντίστροφο

- Επομένως, το άθροισμα των μαζών πιθανότητας όλων των M ακολουθιών που μπορούμε να κωδικοποιήσουμε δεν μπορεί να υπερβαίνει την τιμή

$$\begin{aligned} & 2^{n[H(X)-\zeta]}2^{-n[H(x)-\epsilon]} + \Pr\{X_1^n \notin A_\epsilon^{(n)}\} \\ & = 2^{-n[\zeta-\epsilon]} + \Pr\{X_1^n \notin A_\epsilon^{(n)}\} \\ & \overset{(a)}{<} 2^{-n[\zeta-\epsilon]} + \epsilon. \end{aligned}$$

  (a) από το AEP, για αρκούντως μεγάλο n.
- Συνεπώς, για την πιθανότητα να μην έχουμε κατασκευάσει (δηλαδή να μην υπάρχει διαθέσιμη) κωδική λέξη για μία ακολουθία ισχύει

$$P_e^{(n)} > 1 - 2^{-n[\zeta-\epsilon]} - \epsilon,$$

  για αρκούντως μεγάλο n.

Θεώρημα Κωδικοποίησης Πηγής (5) – αντίστροφο

$$P_e^{(n)} > 1 - 2^{-n[\zeta-\epsilon]} - \epsilon,$$

για αρκούντως μεγάλο n, για οποιοδήποτε $\epsilon > 0$.

- Άρα, ισχύει και για $\epsilon < \zeta$.
- Αλλά για οποιοδήποτε $\epsilon < \zeta$, $P_e > 1 - 2\epsilon$ για αρκούντως μεγάλο n.
- Συνεπώς, $P_e^{(n)} \to 1$ για $n \to \infty$, αφού, για μεγάλο n, και $\epsilon \to 0$.

Θεώρημα Κωδικοποίησης Πηγής (6) – αντίστροφο

- Ένα ερώτημα που προκύπτει εδώ είναι το εξής: Δείξαμε ότι $P_e^{(n)} \to 1$ για $n \to \infty$. Θα μπορούσε κάποιος να ισχυριστεί ότι ίσως να υπάρχει κάποιος τρόπος να κωδικοποιήσουμε με κάποια πεπερασμένη τιμή n και με τον τρόπο αυτό να επιτύχουμε συμπίεση με μέσο μήκος μικρότερο από την εντροπία.
- Μπορούμε να δείξουμε ότι κάτι τέτοιο δεν είναι δυνατό.
  - Έστω ότι υπάρχει τρόπος κωδικοποίησης με κάποιο (σχετικά μικρό) n για τον οποίο $P_e^{(n)} \to 0$.
  - Έστω, τώρα, ότι θέλουμε να κωδικοποιήσουμε μία ακολουθία μήκους Kn, $K \to \infty$. Ένας τρόπος να το επιτύχουμε είναι χωρίζοντάς την σε ακολουθίες μήκους n και χρησιμοποιώντας τη μέθοδο που επιτυγχάνει $P_e^{(n)} \to 0$.
  - Ωστόσο, αυτό σημαίνει ότι βρήκαμε έναν τρόπο να κατασκευάσουμε κώδικα μήκους Kn για τον οποίο $P_e^{(Kn)} \to 0$.
  - Αλλά αυτό είναι άτοπο γιατί δείξαμε ότι, για $n \to \infty$, $P_e^{(n)} \to 1$.

Ασθενές Αντίστροφο Θεώρηματος Κωδικοποίησης Πηγής

- Παρατηρήστε ότι αποδείξαμε όχι μόνο ότι δεν μπορούμε να συμπιέσουμε με ρυθμό μικρότερο από την εντροπία, αλλά και ότι, αν προσπαθήσουμε να συμπιέσουμε με $H(X) - \zeta$, $\zeta > 0$, η πιθανότητα αποτυχίας αποκωδικοποίησης τείνει στο 1.
- Αυτό ονομάζεται *ισχυρό αντίστροφο* (strong converse).
- Θα αποδείξουμε, επίσης, το ασθενές αντίστροφο (weak converse) ότι, δηλαδή, δεν υπάρχει κώδικας με μέσο μήκος μικρότερο από την εντροπία ο οποίος να επιτυγχάνει αυθαίρετα μικρή πιθανότητα αποτυχίας κωδικοποίησης.
- Το ασθενές αντίστροφο προκύπτει από το ισχυρό. Ο λόγος που θα κάνουμε την απόδειξη είναι για να εξοικειωθούμε με τη χρήση της Ανισότητας Fano στην απόδειξη ασθενώς αντιστρόφων.

Ασθενές Αντίστροφο Θεώρηματος Κωδικοποίησης Πηγής (2)

- Έστω ότι κατασκευάζουμε έναν κώδικα συμπίεσης για M ακολουθίες πηγής μήκους n. Επομένως, μπορούμε να γράψουμε $M = 2^{nR}$ όπου R ο μέσος αριθμός των bits ανά σύμβολο πηγής.
- Παρατηρούμε ότι $X_1^n \to M \to \hat{X}_1^n$, όπου X_1^n η ακολουθία που παράγει η πηγή, M ο δείκτης της κωδικής λέξης στο βιβλίο κωδίκων του συμπιεστή (κωδικοποιητή πηγής) και $\hat{X}_1^n$ η αποσυμπιεσμένη ακολουθία στο δέκτη.
- Επομένως, από την Ανισότητα Επεξεργασίας Δεδομένων,

$$
\begin{aligned}
&I(X_1^n; M) \geq I(X_1^n; \hat{X}_1^n) \Rightarrow \\
&H(X_1^n) - H(X_1^n|M) \geq H(X_1^n) - H(X_1^n|\hat{X}_1^n) \Rightarrow \\
&H(X_1^n|M) \leq H(X_1^n|\hat{X}_1^n).
\end{aligned}
$$

Ασθενές Αντίστροφο Θεώρηματος Κωδικοποίησης Πηγής (3)

- Από την Ανισότητα Fano,

$$\begin{aligned} H(X_1^n|M) \leq H(X_1^n|\hat{X}_1^n) &\leq nP_e^{(n)} \log|\mathcal{X}| + 1 \\ &= n\left(P_e^{(n)} \log|\mathcal{X}| + \frac{1}{n}\right) \triangleq n\epsilon_n. \end{aligned}$$

- Επειδή θέλουμε ο κώδικας να επιτυγχάνει $P_e^{(n)} \to 0$ για $n \to \infty$, $\epsilon_n \to 0$ για $n \to \infty$.

Ασθενές Αντίστροφο Θεώρηματος Κωδικοποίησης Πηγής (4)

- Επομένως

$$\begin{aligned} nR &\overset{(a)}{\geq} H(M) \overset{(b)}{=} H(M) - H(M|X_1^n) \\ &= I(M; X_1^n) = H(X_1^n) - H(X_1^n|M) \\ &\overset{(c)}{\geq} nH(X) - n\epsilon_n. \end{aligned}$$

  (a) $M = 2^{nR}$. (b) Ο δείκτης, M, της κωδικής λέξης είναι ντετερμινιστική συνάρτηση της ακολουθίας X_1^n της πηγής. (c) Από την Ανισότητα Fano.

- Συνεπώς, για $n \to \infty$, $R \geq H(X)$.
- Η Ανισότητα Fano είναι ιδιαίτερα χρήσιμη στην απόδειξη του ασθενούς αντιστρόφου. Θα την χρησιμοποιήσουμε ξανά στα κανάλια.

Θεώρημα Κωδικοποίησης Πηγής (7)

- Επομένως, αποδείξαμε και το αντίστροφο του θεωρήματος Κωδικοποίησης Πηγής, ότι, δηλαδή, δεν μπορεί να επιτευχθεί συμπίεση χωρίς απώλειες με μέσο μήκος μικρότερο της εντροπίας.
- Το Θεώρημα Κωδικοποίησης Πηγής για κωδικοποίηση μεταβλητού μήκους είναι πιο "ισχυρό" από το Θεώρημα Κωδικοποίησης Πηγής για κωδικοποίηση σταθερού μήκους, δεδομένου ότι στο όριο η συμπίεση μεταβλητού μήκους συμπίπτει με τη συμπίεση σταθερού μήκους.
- Το Θεώρημα Κωδικοποίησης Πηγής ισχύει και για διακριτές στάσιμες εργοδικές πηγές με $H(X) < \infty$: Μπορούμε να συμπιέσουμε με μέσο μήκος που τείνει στο ρυθμό εντροπίας $H(\mathcal{X})$. Ωστόσο, η απόδειξη είναι πιο πολύπλοκη (βλ. π.χ. Gallager 3.5.)
- Στα επόμενα θα θεωρούμε ότι η μέγιστη συμπίεση χωρίς απώλειες που μπορεί να επιτευχθεί ισούται με το ρυθμό εντροπίας (ο οποίος, για πηγές χωρίς μνήμη, ταυτίζεται με την εντροπία ανά σύμβολο).

Αποδοτική Κωδικοποίηση Πηγής

- Έστω ότι, με χρήση κώδικα, η ακολουθία $(X_1, X_2, \ldots, X_n)$ μίας πηγής κωδικοποιείται στη *δυαδική* ακολουθία $(Y_1, Y_2, \ldots, Y_m)$. Θεωρούμε ότι οι X_i είναι i.i.d. (όχι, απαραίτητα, δυαδικές), δηλαδή ότι η πηγή δεν έχει μνήμη.
- Έστω, επίσης, ότι το αλφάβητο $\mathcal{X}$ της πηγής είναι πεπερασμένο (για απλοποίηση).
- Από το AEP, για $n \to \infty$, $m \approx nH(X)$.
- Εάν $\hat{X}_1^n$ είναι η ανακατασκευασμένη (αποσυμπιεσμένη) ακολουθία, η πιθανότητα εσφαλμένης αποκωδικοποίησης είναι $P_e = \Pr\{X_1^n \neq \hat{X}_1^n\}$.
- Θα δείξουμε ότι, εάν απαιτήσουμε $P_e \to 0$ για $n \to \infty$, τα σύμβολα Y_i της ακολουθίας Y_1^m είναι (σχεδόν) i.i.d. Bern$(1/2)$.

Αποδοτική Κωδικοποίηση Πηγής (2)

- Από την ανισότητα Fano,

$$H(X_1^n|\hat{X}_1^n) \leq 1 + P_e \log |\mathcal{X}|^n = 1 + nP_e \log |\mathcal{X}|.$$

- Επειδή $\hat{X}_1^n = f(Y_1^m)$, $H(Y_1^m) = H(Y_1^m, \hat{X}_1^n) \geq H(\hat{X}_1^n)$.
- Επομένως,

$$\begin{aligned} H(Y_1^m) &\geq H(\hat{X}_1^n) \geq H(\hat{X}_1^n) - H(\hat{X}_1^n|X_1^n) \\ &= I(X_1^n; \hat{X}_1^n) = H(X_1^n) - H(X_1^n|\hat{X}_1^n) \\ &= nH(X) - H(X_1^n|\hat{X}_1^n) \\ &\overset{(a)}{\geq} nH(X) - (1 + nP_e \log |\mathcal{X}|) \\ &= n(H(X) - P_e \log |\mathcal{X}|) - 1. \end{aligned}$$

(a) Ανισότητα Fano.

Αποδοτική Κωδικοποίηση Πηγής (3)

$$H(Y_1^m) \geq n(H(X) - P_e \log |\mathcal{X}|) - 1.$$

- Επίσης, από το φράγμα ανεξαρτησίας της εντροπίας,

$$H(Y_1^m) \leq \sum_{i=1}^{m} H(Y_i) \leq m,$$

  επειδή έχουμε υποθέσει ότι οι Y_i είναι δυαδικές.
- Συνεπώς,

$$n(H(X) - P_e \log |\mathcal{X}|) - 1 \leq H(Y_1^m) \leq m.$$

- Αλλά για $P_e \to 0$ και $n \to \infty$, το κάτω φράγμα τείνει στο $nH(X) \approx m$.

Αποδοτική Κωδικοποίηση Πηγής (4)

- Επομένως, $H(Y_1^m) \approx m$.
- Δηλαδή, η ακολουθία Y_1^m έχει τη μέγιστη δυνατή εντροπία (είναι όσο πιο τυχαία γίνεται).
- Διαισθητικά, αν η $H(Y_1^m)$ δεν ήταν εντελώς τυχαία, θα μπορούσαμε να τη συμπιέσουμε περισσότερο, οπότε ο τρόπος που χρησιμοποιήσαμε αρχικά δε θα ήταν βέλτιστος.