EE728
Προχωρημένα Θέματα Θεωρίας Πληροφορίας
3η διάλεξη

Δημήτρης-Αλέξανδρος Τουμπακάρης

Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών

3 Μαρτίου 2010

Περιεχόμενα σημερινού μαθήματος

1. Η ιδιότητα ασυμπτωτικής ισοδιαμέρισης (AEP)
   - Ασθενής Τυπικότητα
   - Ισχυρή Τυπικότητα
2. Κωδικοποίηση Σταθερού Μήκους και Θεώρημα Κωδικοποίησης Πηγής
   - Κωδικοποίηση Σταθερού Μήκους
   - Θεώρημα Κωδικοποίησης Πηγής
3. Ανισότητα επεξεργασίας δεδομένων και Ανισότητα Fano
   - Ανισότητα επεξεργασίας δεδομένων
   - Ανισότητα Fano

Αντιστοιχία με βιβλία Cover & Thomas και Gallager

- Cover & Thomas: 2.8, 2.10, 3.1–3.3
- Gallager: 3.1, 4.3 (έως Theorem 4.3.3)

Τυπικό Σύνολο (Typical Set) και ιδιότητες

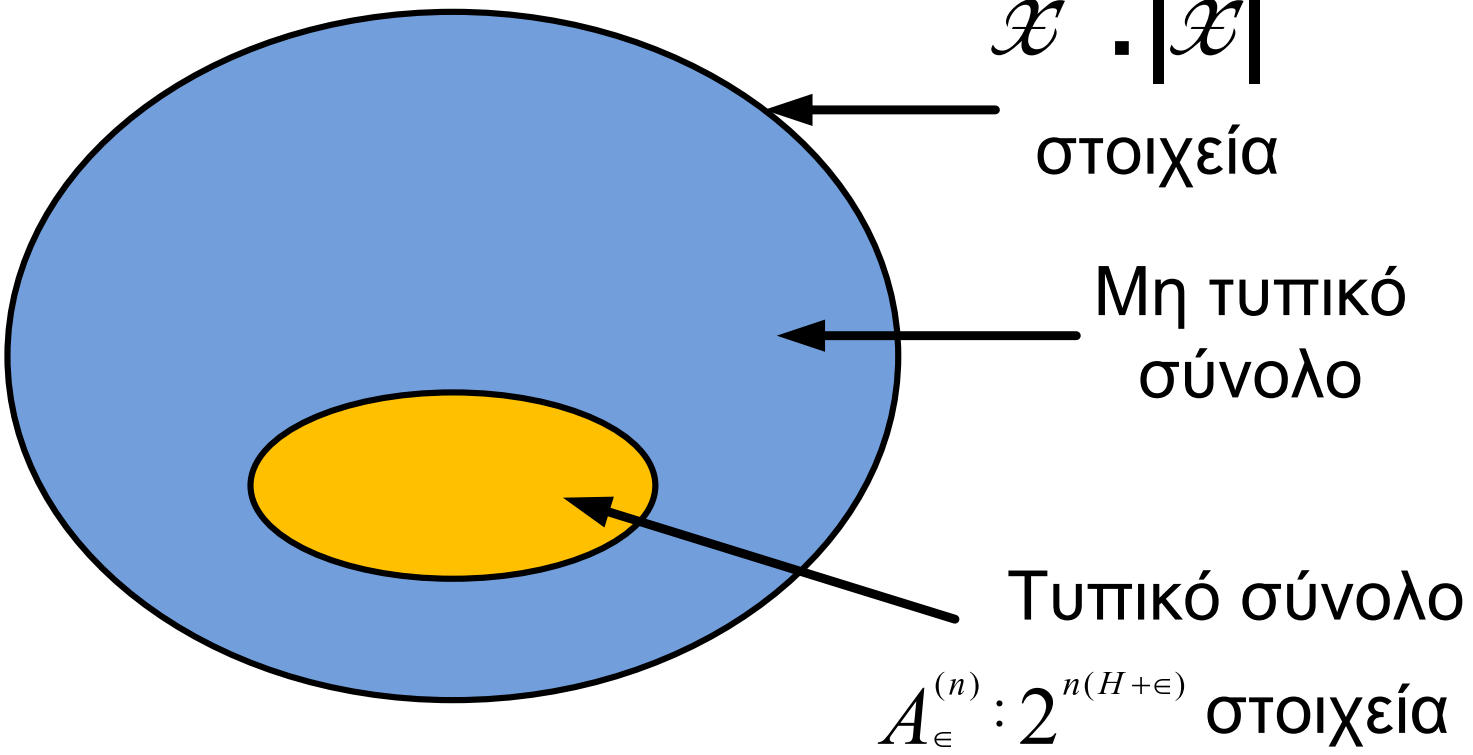
- Το (ασθενώς) τυπικό σύνολο $A_\epsilon^{(n)}$ που αντιστοιχεί στην κατανομή $p(x)$ αποτελείται από τις ακολουθίες $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ που ικανοποιούν τη σχέση

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

- 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 $n \geq n_0$,

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

Παράδειγμα 3.1 (Cover Problem 3.6)

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

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

- Απάντηση:

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

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

Σχέση τυπικού συνόλου με σύνολα που περιέχουν σχεδόν όλη την πιθανότητα

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

Ισχυρή Τυπικότητα (Strong Typicality)

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

Ισχυρώς Τυπικό Σύνολο – ορισμός

- Θεωρούμε πηγή χωρίς μνήμη με κατανομή $p(x)$. Έστω ότι $\mathcal{S}_X \subseteq \mathcal{X}$ είναι το σύνολο στο οποίο $p(x) > 0$.
- Το ισχυρώς τυπικό σύνολο $T^n_{[X]\delta}$ 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

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

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

- Οι ακολουθίες που ανήκουν στο $T^n_{[X]\delta}$ ονομάζονται ισχυρώς δ-τυπικές.

Ισχυρή Τυπικότητα – σχόλια

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

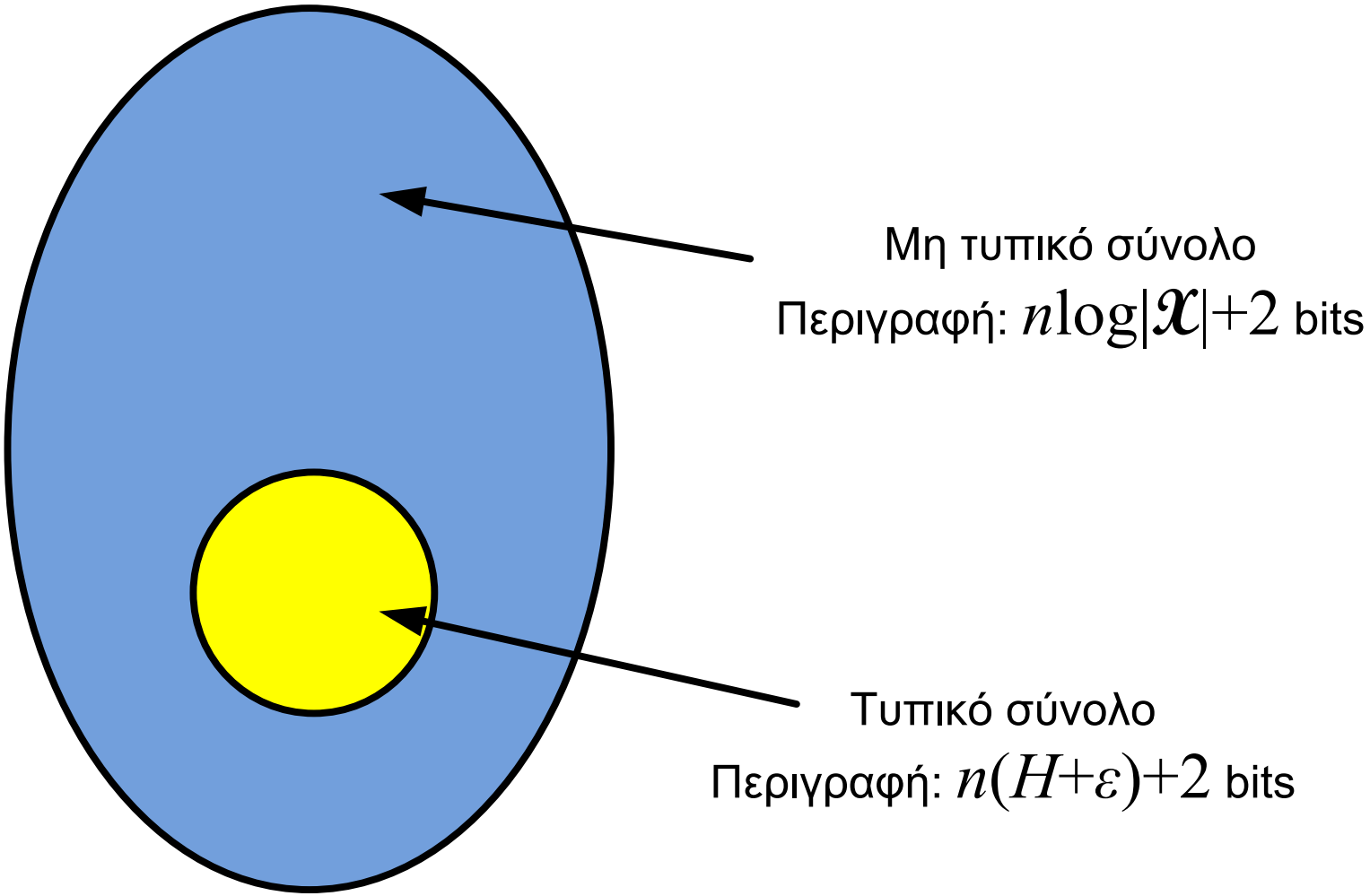
Κωδικοποίηση Σταθερού Μήκους και Θεώρημα Κωδικοποίησης Πηγής

1. Η ιδιότητα ασυμπτωτικής ισοδιαμέρισης (AEP)
   - Ασθενής Τυπικότητα
   - Ισχυρή Τυπικότητα
2. Κωδικοποίηση Σταθερού Μήκους και Θεώρημα Κωδικοποίησης Πηγής
   - Κωδικοποίηση Σταθερού Μήκους
   - Θεώρημα Κωδικοποίησης Πηγής
3. Ανισότητα επεξεργασίας δεδομένων και Ανισότητα Fano
   - Ανισότητα επεξεργασίας δεδομένων
   - Ανισότητα Fano

Κωδικοποίηση Σταθερού Μήκους

- Έστω, όπως και προηγουμένως, 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 $A_\epsilon^{(n)^c} = \mathcal{X}^n - A_\epsilon^{(n)}$.
- Διατάσσουμε όλες τις ακολουθίες σε κάθε σύνολο. Για το τυπικό σύνολο, δεδομένου ότι περιέχει το πολύ $2^{n(H+\epsilon)}$ ακολουθίες (σύμφωνα με την Ιδιότητα 3), χρειαζόμαστε το πολύ $n(H+\epsilon)+1$ bits (το επιπλέον 1 bit οφείλεται στο ότι ενδέχεται η ποσότητα $n(H+\epsilon)$ να μην είναι ακέραιος).
- Για το μη τυπικό σύνολο, χρειαζόμαστε το πολύ $n \log |\mathcal{X}| + 1$ bits.
- 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, αλλιώς χρησιμοποιούμε πρόθεμα 1.

Κωδικοποίηση Σταθερού Μήκους με χρήση τυπικού συνόλου



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 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 αποδοτικότερη (χρησιμοποιώντας, π.χ. $n \log \left| A_\epsilon^{(n)^c} \right|$ bits).
- Παρατηρήστε ότι το τυπικό σύνολο ενδέχεται να περιέχει λίγα στοιχεία (το μέγεθός του είναι $\sim 2^{nH}$). 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

- 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, $H(X) \leq \mathbb{E}[l^*] < H(X) + 1 \Rightarrow H(X^L) \leq \mathbb{E}[\tilde{l}^*] < H(X^L) + 1 \Rightarrow LH(X) \leq \mathbb{E}[\tilde{l}^*] < LH(X) + 1 \Rightarrow H(X) \leq \mathbb{E}[\tilde{l}^*]/L < H(X) + 1/L$.
- Επομένως, υπάρχει και δεύτερος τρόπος να συμπιέσουμε κοντά στην εντροπία, αυτή τη φορά με κώδικα μεταβλητού μήκους.
- Απομένει να αποδείξουμε ότι, εάν προσπαθήσουμε να συμπιέσουμε με μέσο μήκος μικρότερο από την εντροπία, η πιθανότητα σφάλματος $P_e \to 1$.

Θεώρημα Κωδικοποίησης Πηγής (2)

- Έστω ότι το μήκος της αρχικής (προς συμπίεση) ακολουθίας ισούται με L. Θεωρούμε δυαδικές ακολουθίες (αν και η απόδειξη γενικεύεται εύκολα). Έστω ότι η ακολουθία συμπιέζεται με χρήση N bits, όπου $N < L[H(X) - 2\epsilon]$, $\epsilon > 0$. Επομένως, μπορούμε να ανακατασκευάσουμε το πολύ $2^{L(H(X)-2\epsilon)}$ ακολουθίες στην έξοδο του αποκωδικοποιητή.
- Ας υποθέσουμε, κατ' αρχάς, ότι κωδικοποιούμε μόνο τυπικές ακολουθίες.
- Η πιθανότητα P_c να μπορέσουμε να κωδικοποιήσουμε μια ακολουθία επιτυχώς ισούται με την πιθανότητα να εμφανιστεί μία από τις $2^{L(H(X)-2\epsilon)}$ τυπικές ακολουθίες που μπορούμε να κωδικοποιήσουμε.

Θεώρημα Κωδικοποίησης Πηγής (3)

- Δεδομένου ότι η από κοινού συνάρτηση μάζας πιθανότητας μιας τυπικής ακολουθίας δεν υπερβαίνει την τιμή $2^{-L(H(X)-\epsilon)}$, $P_c \leq 2^{-L(H(X)-\epsilon)} \cdot 2^{L(H(X)-2\epsilon)} = 2^{-L\epsilon}$.
- Συνεπώς, για την πιθανότητα αποτυχίας κωδικοποίησης και, επομένως, σφάλματος, ισχύει $P_e = 1 - P_c \geq 1 - 2^{-L\epsilon}$. Για $L \to \infty$, $P_e \to 1$ για οποιοδήποτε $\epsilon > 0$.
- Αν χρησιμοποιήσουμε κάποιες από τις διαθέσιμες κωδικές λέξεις για να κωδικοποιήσουμε μη τυπικές ακολουθίες, ισχύει $P_c \leq 2^{-L(H(X)-\epsilon)} \cdot 2^{L(H(X)-2\epsilon)} + \zeta$, όπου $\zeta \to 0$ καθώς $L \to \infty$ (από το AEP, γιατί η πιθανότητα να εμφανιστεί μη τυπική ακολουθία τείνει στο 0).
- Συνεπώς, για $L \to \infty$, $P_e \to 1$ για οποιοδήποτε $\epsilon > 0$.

Θεώρημα Κωδικοποίησης Πηγής (4)

- 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 ταυτίζεται με την εντροπία ανά σύμβολο για πηγές χωρίς μνήμη).

Ανισότητα επεξεργασίας δεδομένων και Ανισότητα Fano

1. Η ιδιότητα ασυμπτωτικής ισοδιαμέρισης (AEP)
   - Ασθενής Τυπικότητα
   - Ισχυρή Τυπικότητα
2. Κωδικοποίηση Σταθερού Μήκους και Θεώρημα Κωδικοποίησης Πηγής
   - Κωδικοποίηση Σταθερού Μήκους
   - Θεώρημα Κωδικοποίησης Πηγής
3. Ανισότητα επεξεργασίας δεδομένων και Ανισότητα Fano
   - Ανισότητα επεξεργασίας δεδομένων
   - 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).
- $X \to Y \to g(Y)$.
- Ανισότητα Επεξεργασίας Δεδομένων (Data Processing Inequality): Εάν $X \to Y \to Z$, τότε $I(X; Y) \geq I(X; Z)$.

Ανισότητα Επεξεργασίας Δεδομένων (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 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 οι αμοιβαίες πληροφορίες 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)

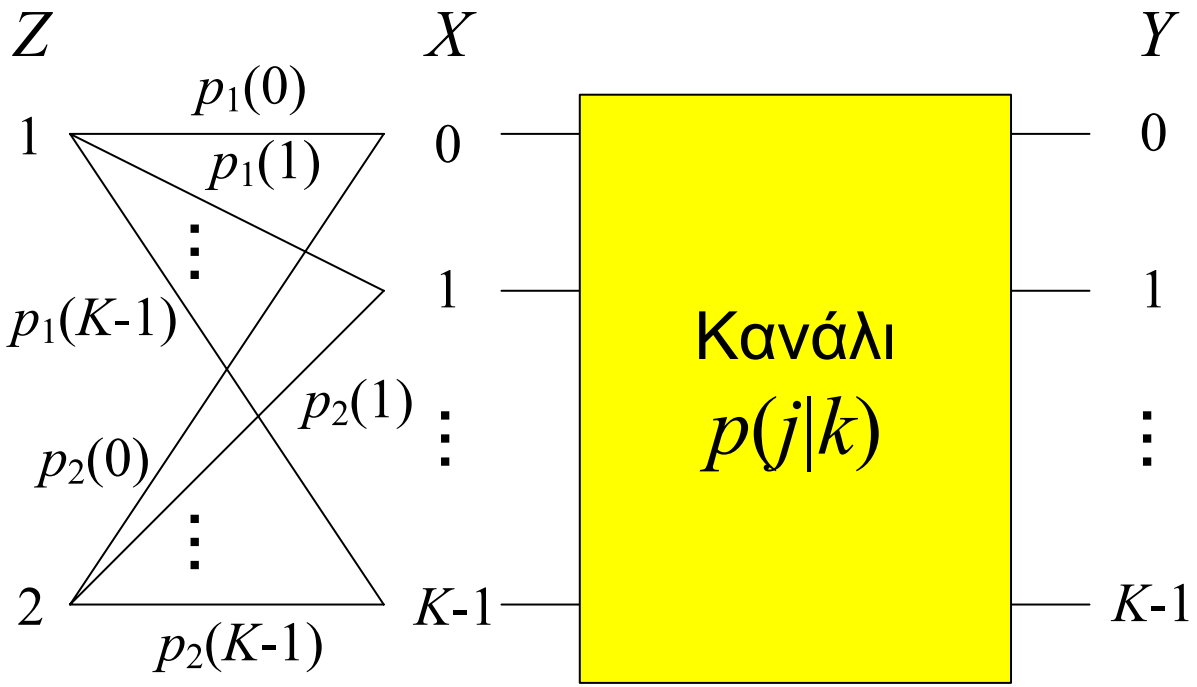
- Μπορούμε να υποθέσουμε ότι οι p_1 και p_2 είναι υπό συνθήκη κατανομές που εξαρτώνται από μια δυαδική τ.μ. Z:

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

- 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, $I(Y;Z|X) = 0$.
- 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 αποδειχθεί ότι η $I(X;Y)$ είναι κυρτή ($\cup$) συνάρτηση της $p(y|x)$ για δεδομένη $p(x)$ (Gallager Theorem 4.4.3).

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 νομοτελειακός (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 $P_e = \Pr\{\hat{X} \neq X\}$.

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, νομοτελειακή 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 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 $P_e, H(X|Y)$

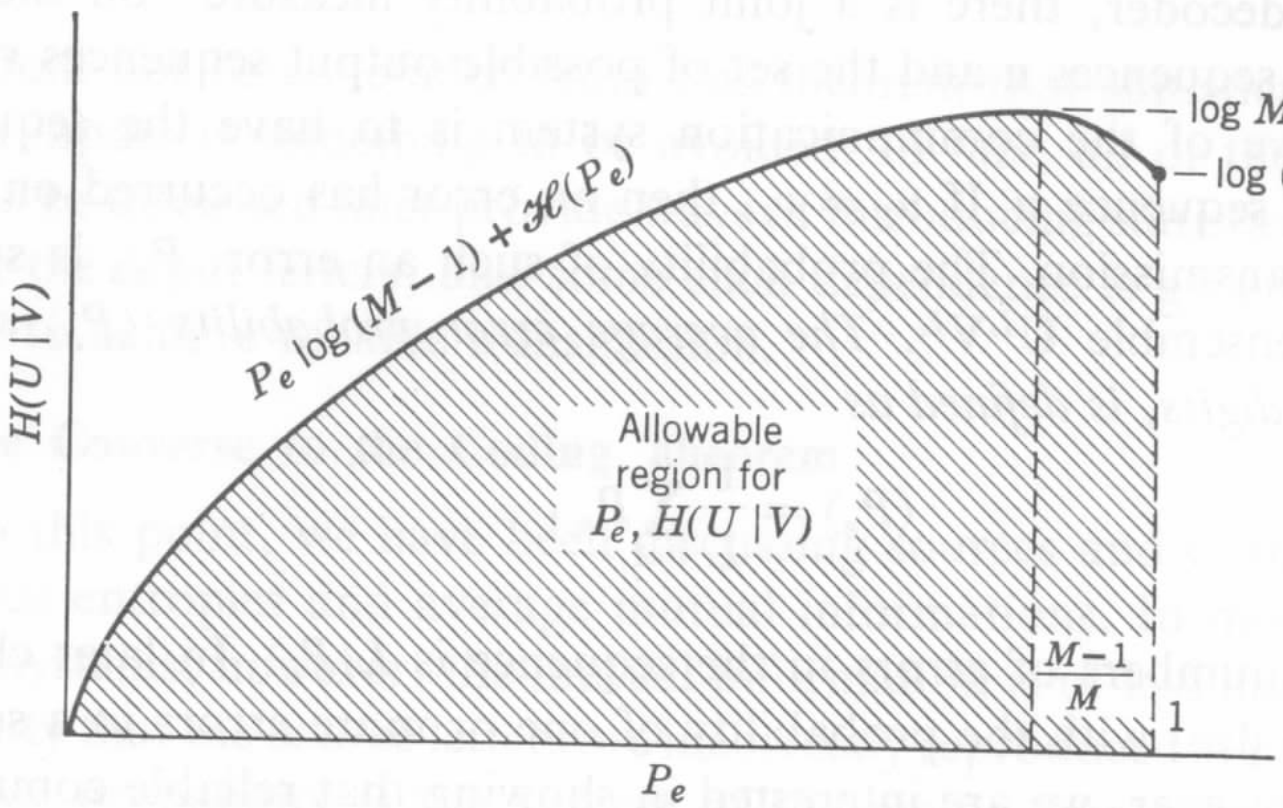


Figure 4.3.2. Interpretation of Theorem 4.3.1.

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