EE728
Προχωρημένα Θέματα Θεωρίας Πληροφορίας
2η διάλεξη
(3η έκδοση, 11/3)

Δημήτρης-Αλέξανδρος Τουμπακάρης

Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών

19 Φεβρουαρίου 2013

Περιεχόμενα 2ης διάλεξης

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

- Βιβλίο Cover & Thomas (2η έκδοση): Κεφ. 3, Κεφ. 2.1 – 2.5
- Βιβλίο El Gamal & Kim: Κεφ. 2.4, 2.1, 2.3

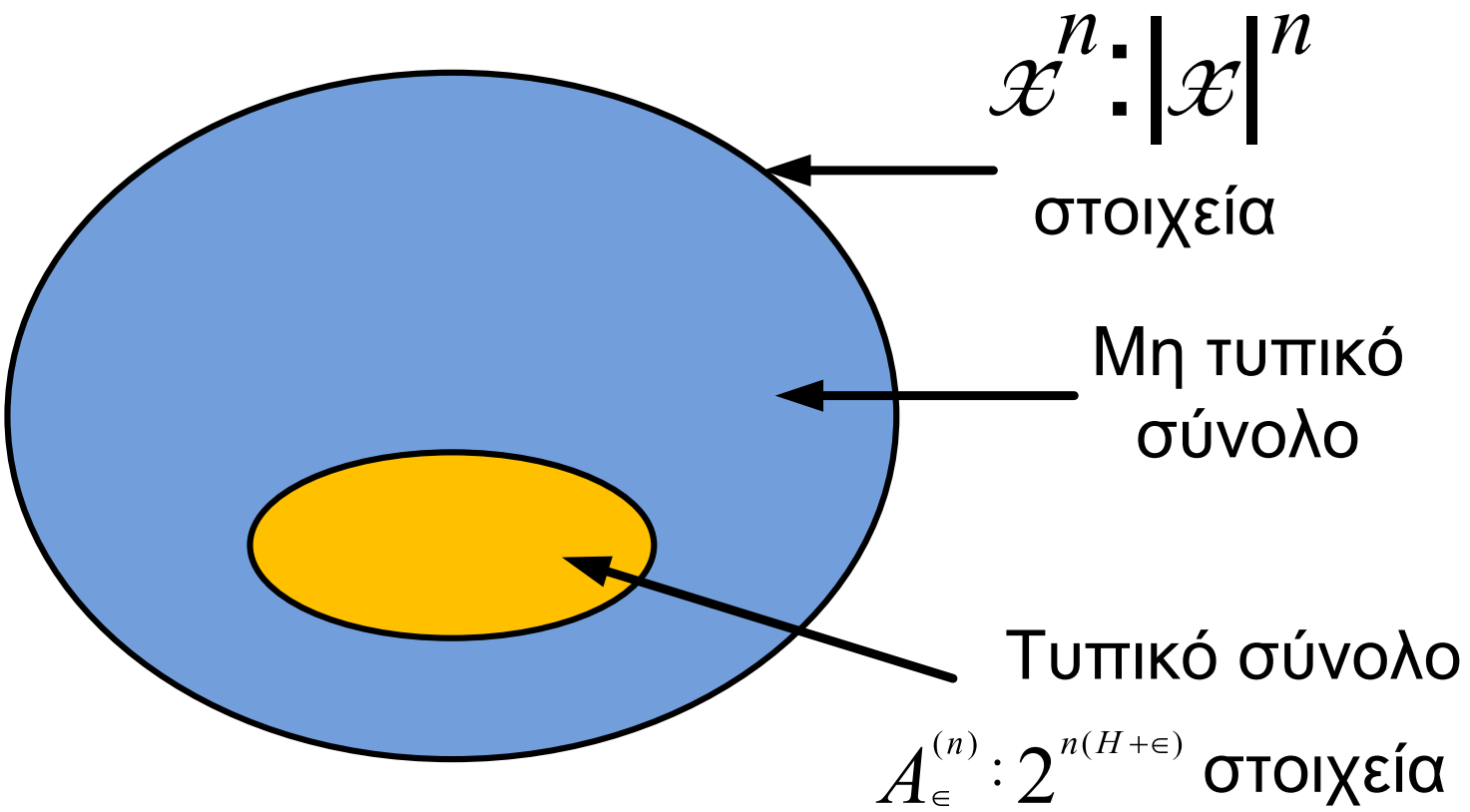
Τυπικό Σύνολο (Typical Set) και ιδιότητες

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

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

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

Τυπικό Σύνολο



Αποδείξεις ιδιοτήτων Τυπικού Συνόλου

1. Εάν $(x_1, x_2, \ldots, x_n) \in A_\epsilon^{(n)}$,
   $H(X) - \epsilon \leq -\frac{1}{n} \log p(x_1, x_2, \ldots, x_n) \leq H(X) + \epsilon$.
   Προκύπτει άμεσα από τον ορισμό του 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.

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

Από τη 2η ιδιότητα, για $n \geq n_0$,

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

Αποδείξεις ιδιοτήτων Τυπικού Συνόλου (3)

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

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

Παράδειγμα 2.1 (Cover & Thomas Problem 3.6)

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

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

- Απάντηση:

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

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

Σχέση τυπικού συνόλου με σύνολα που περιέχουν σχεδόν όλη την πιθανότητα

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

Ισχυρή Τυπικότητα (Strong Typicality)

- Έως τώρα ασχοληθήκαμε με την ασθενή τυπικότητα.
- Μια ακολουθία είναι ασθενώς τυπική όταν η εμπειρική της εντροπία βρίσκεται κοντά στην πραγματική εντροπία της πηγής που παράγει την ακολουθία.
- Για να είναι μια ακολουθία ισχυρώς τυπική πρέπει η σχετική συχνότητα με την οποία εμφανίζεται κάθε σύμβολο μέσα στην ακολουθία να βρίσκεται κοντά στην κατανομή της πηγής.
- Για παράδειγμα, για πηγή Bern$(1/2)$, η ακολουθία $0\ 0\ 0\ 1\ 0\ 0\ 0$ είναι ασθενώς τυπική, αλλά όχι ισχυρώς τυπική (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.
- 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 ($\log \to \log_2$). Σπανιότερα, nat ($\log \to \ln$).
- $H_b(X) = \log_b a \cdot H_a(X)$.
- Από εδώ και στο εξής $\log$ υπονοεί $\log_2$ (αν και δεν έχει ιδιαίτερη σημασία ποια μονάδα χρησιμοποιούμε).

Από κοινού και υπό συνθήκη εντροπία

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

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

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

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

Ιδιότητες Εντροπίας διακριτής τ.μ.

- $H(X) \geq 0$.
- Η εντροπία είναι κοίλη ($\cap$) συνάρτηση της συνάρτησης μάζας πιθανότητας $p(x)$. Θα το αποδείξουμε.
- $H(X) \leq \log |\mathcal{X}|$, όπου $|\mathcal{X}|$ το μέγεθος του αλφαβήτου της X. Το μέγιστο επιτυγχάνεται από την ομοιόμορφη κατανομή: $p(X_i) = \frac{1}{|\mathcal{X}|}$ για όλα τα $X_i \in \mathcal{X}$. 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 $H(\mathcal{X}) = H'(\mathcal{X})$.

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.