ΕΕ728

Προχωρημένα Θέματα Θεωρίας Πληροφορίας

Δημήτρης - Αλέξανδρος Τουμπακάρης

2ο Μάθημα – 13 Μαρτίου 2009

Ανακεφαλαίωση προηγούμενου μαθήματος

- Επανάληψη Βασικών Ποσοτήτων Θεωρίας Πληροφορίας και ιδιοτήτων τους
  - Εντροπία (διακριτής τ.μ.)
  - Δεσμευμένη Εντροπία
  - Από κοινού Εντροπία

Προεπισκόπηση σημερινού μαθήματος

- Συνέχεια Επανάληψης
  - Σχετική Εντροπία
  - Αμοιβαία Πληροφορία.
  - Κυρτές συναρτήσεις και ανισότητα Jensen.
  - Ιδιότητες Εντροπίας, Σχετικής Εντροπίας και Αμοιβαίας Πληροφορίας.

- Ανισότητα Επεξεργασίας Δεδομένων.

- Ανισότητα Fano.

- Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP): Η "καρδιά" της συμπίεσης (και της Θεωρίας Πληροφορίας).

Ανισότητα Επεξεργασίας Δεδομένων

- Επανάληψη Βασικών Ποσοτήτων Θεωρίας Πληροφορίας (συνέχεια)
- Ανισότητα Επεξεργασίας Δεδομένων
- Ανισότητα Fano
- Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP)

Σχετική Εντροπία $D(p||q)$

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

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

- Προσοχή: Η μέση τιμή είναι ως προς την κατανομή p.

- Από πού πηγάζει αυτός ο ορισμός; Όπως είδαμε στη ‟Θεωρία Πληροφορίας´´, η $D(p||q)$ ποσοτικοποιεί τα επιπλέον bits που χρειαζόμαστε για να συμπιέσουμε μια τ.μ. με πραγματική κατανομή p όταν για τη συμπίεση χρησιμοποιείται η κατανομή q.

Σχετική Εντροπία $D(p||q)$ (συνέχεια)

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

- $D(p||q) \geq 0$. Αποδείχθηκε στη "Θεωρία Πληροφορίας" με χρήση της ανισότητας **Jensen** και του γεγονότος ότι η log είναι κοίλη ($\cap$). Ωστόσο, η $D(p||q)$ δεν είναι απόσταση κατά την αυστηρή έννοια: $D(p||q) \neq D(q||p)$. Επίσης, δεν ισχύει η τριγωνική ανισότητα.

Δεσμευμένη Σχετική Εντροπία και Κανόνας Αλυσίδας

- Δεσμευμένη σχετική εντροπία (conditional relative entropy):

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

- Κανόνας αλυσίδας για τη σχετική εντροπία

$$D(p(x,y)||q(x,y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x)).$$

- Απόδειξη: Απλή, με χρήση ορισμού (Cover Theorem 2.5.3).

Αμοιβαία Πληροφορία $I(X;Y)$

- Έστω μια τ.μ. $X \sim p(X)$. Εάν μας γνωστοποιηθεί η τιμή της τ.μ. Y, η κατανομή πιθανότητας της X αλλάζει σε $p(X|Y)$. Επομένως, κατά μέσο όρο, γνώση της Y αλλάζει την αβεβαιότητα που έχουμε για τη X κατά $E_p\left[\frac{p(X|Y)}{p(X)}\right]$ (η μέση τιμή υπολογίζεται για όλες τις τιμές των X και Y).

- Συνεπώς,

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

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 <u>κατά μέσο όρο</u>.

- 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 <u>κατά μέσο όρο</u>.

- Προσοχή: Για κάποιες τιμές της Y, ενδέχεται $I(X;Y=y) < 0$. Ωστόσο, ισχύει πάντα $I(X;Y) = E_{p(Y)}[I(X;Y=y)] \geq 0$.

Αμοιβαία Πληροφορία $I(X;Y)$ (3)

- Μια διαφορετική ερμηνεία της αμοιβαίας πληροφορίας με βάση τη σχετική εντροπία: Η πληροφορία που "χάνουμε" εάν θεωρήσουμε ότι οι X και Y είναι ανεξάρτητες, ενώ, στην πραγματικότητα, δεν είναι.

- $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$. Προκύπτει από τον ορισμό (αποδείχθηκε στη "Θεωρία Πληροφορίας").

Αμοιβαία Πληροφορία $I(X;Y)$ (4)

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

- Κανόνας αλυσίδας για την αμοιβαία πληροφορία:

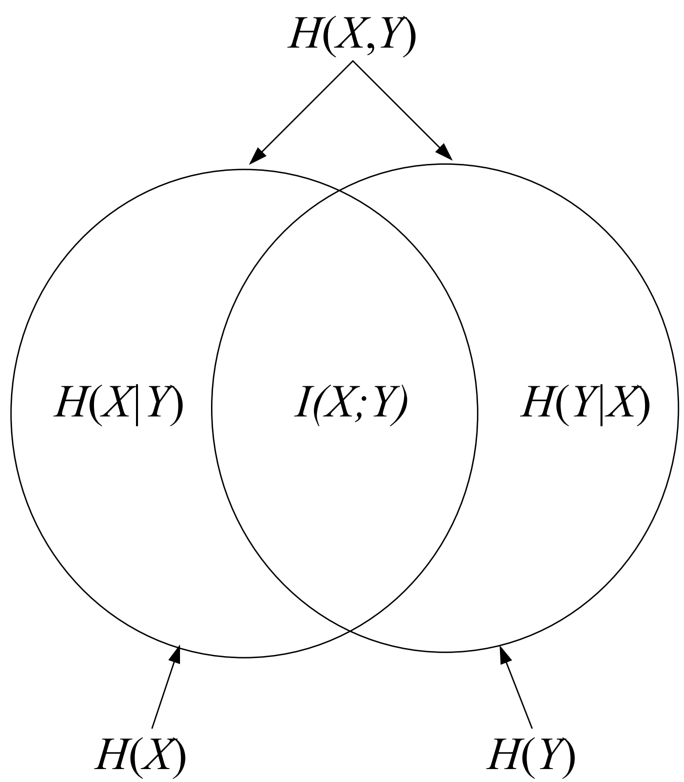
$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y | X_1, X_2, \ldots, X_{i-1}).$$

- Απόδειξη: Εύκολα, από κανόνα αλυσίδας εντροπίας και χρήση $I(X_1, X_2, \ldots, X_n; Y) = H(X_1, X_2, \ldots, X_n) - H(X_1, X_2, \ldots, X_n|Y)$.

- Υπό συνθήκη αμοιβαία πληροφορία: $I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$.

Διάγραμμα **Venn**

Η σχέση μεταξύ εντροπίας, δεσμευμένης εντροπίας και αμοιβαίας πληροφορίας μπορεί να αναπαρασταθεί και με χρήση διαγράμματος Venn.



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

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

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

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

- Πρακτικά, μια συνάρτηση είναι κυρτή όταν μια χορδή που ενώνει δύο οποιεσδήποτε τιμές της δε βρίσκεται ποτέ "κάτω" από τη συνάρτηση.

- Παραδείγματα κυρτών συναρτήσεων: x^2, $|x|$, e^x, $x \log x$ (για $x \geq 0$).

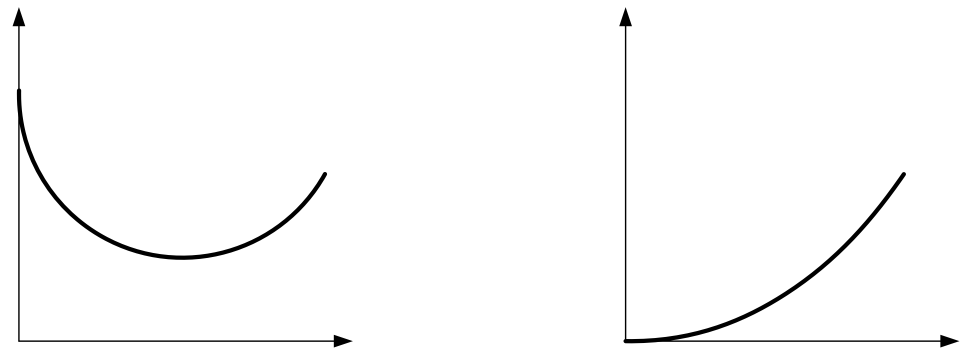
Κυρτές (**convex**) και κοίλες (**concave**) συναρτήσεις (συνέχεια)

- Ορισμός: Μια συνάρτηση $f(x)$ είναι (αυστηρώς) κοίλη ($\cap$) σε διάστημα (a, b) εάν η $-f(x)$ είναι (αυστηρώς) κυρτή.

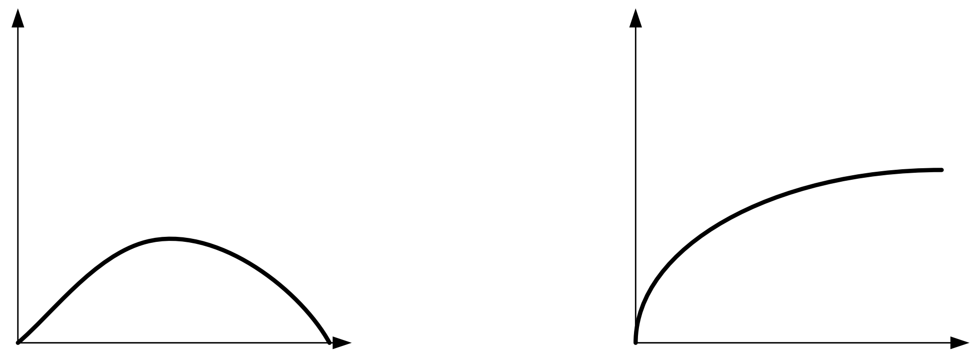
- Παραδείγματα κοίλων συναρτήσεων: $\log x$, $\sqrt{x}$ (για $x \geq 0$).

- Η συνάρτηση $ax + b$ (affine) είναι κυρτή και κοίλη.

Παραδείγματα κυρτών και κοίλων συναρτήσεων



(α) Κυρτές συναρτήσεις



(β) Κοίλες συναρτήσεις

Ανισότητα **Jensen**

- Θεώρημα: Μια διαφορίσιμη συνάρτηση είναι (αυστηρώς) κυρτή ($\cup$) σε ένα διάστημα όταν έχει μη αρνητική (θετική) δεύτερη παράγωγο στο διάστημα αυτό.

- Απόδειξη: Σε βιβλία ανάλυσης ή Cover Theorem 2.6.1

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

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

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

Απόδειξη ανισότητας **Jensen**

- Για τ.μ. με δύο ενδεχόμενα, από τον ορισμό της κυρτότητας, $p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 x_1 + p_2 x_2)$ (δεδομένου ότι $p_2 = 1 - p_1$).
- Έστω ότι η σχέση ισχύει για τ.μ. με $k-1$ ενδεχόμενα.
- Θέτουμε $p_i' = \frac{p_i}{1-p_k}$, για $i = 1, 2, \ldots, k-1$.

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

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 $\mathcal{A} = \{x : p(x) > 0\}$.

$$-D(p||q) = -\sum_{x\in\mathcal{A}} p(x)\log\frac{p(x)}{q(x)} = \sum_{x\in\mathcal{A}} p(x)\log\frac{q(x)}{p(x)} =$$

$$\stackrel{(a)}{\leq} \log\sum_{x\in\mathcal{A}} p(x)\frac{q(x)}{p(x)} = \log\sum_{x\in\mathcal{A}} q(x) \stackrel{(b)}{\leq} \log\sum_{x\in\mathcal{X}} q(x) = \log 1 = 0.$$

- Στο (a) χρησιμοποιήθηκε το γεγονός ότι η $\log t$ είναι αυστηρώς κοίλη συνάρτηση του t. (b) γιατί;
- Η ισότητα ισχύει εάν και μόνο εάν $q(x)/p(x) = c$ για όλα τα x, δηλαδή εάν $q(x) = cp(x)$. Επίσης, πρέπει $\sum_{x\in\mathcal{A}} q(x) = \sum_{x\in\mathcal{X}} q(x) = \sum_{x\in\mathcal{X}} cp(x) = c = 1$. Συνεπώς, $D(p||q) = 0 \Leftrightarrow p(x) = q(x)$ για όλα τα $x \in \mathcal{A}$.

Συνέπειες ανισότητας πληροφορίας

- Η αμοιβαία πληροφορία είναι πάντοτε μη αρνητική: Για οποιεσδήποτε τ.μ. X και Y, $\boxed{I(X;Y) \geq 0}$. Προκύπτει άμεσα από τον ορισμό της $I(X;Y)$ και από την ανισότητα πληροφορίας.

- $D(p(y|x)||q(y|x)) \geq 0$ (Γιατί; Πότε ισχύει η ισότητα;)

- $I(X;Y|Z) \geq 0$.

- $H(X|Y) \leq H(X)$. Δεδομένου ότι $I(X;Y) \geq 0 \Rightarrow H(X) - H(X|Y) \geq 0$.

- Προσοχή: Δεν ισχύει πάντα $H(X|Y = y) \leq H(X)$ (και, επομένως, δεν ισχύει πάντα ότι $I(X;Y = y) \geq 0$).

Φράγμα Ανεξαρτησίας (**Independence Bound**) Από Κοινού Εντροπίας

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

- Η ισότητα ισχύει εάν και μόνο εάν οι X_i είναι ανεξάρτητες.

Άνω φράγμα $H(X)$ δεδομένου του πλήθους ενδεχομένων $|\mathcal{X}|$

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

- Έστω $u(x) = \frac{1}{|\mathcal{X}|}$ η (διακριτή) ομοιόμορφη κατανομή μάζας πιθανότητας στο σύνολο $\mathcal{X}$ και $p(x)$ η κατανομή μάζας πιθανότητας της X. Από τον ορισμό της σχετικής εντροπίας, $D(p||u) = \sum p(x) \log \frac{p(x)}{u(x)} = \log|\mathcal{X}| - H(X)$.

- Από την ανισότητα πληροφορίας, $0 \leq D(p||u) = \log|\mathcal{X}| - H(X) \Rightarrow H(X) \leq \log|\mathcal{X}|$.

- Η ισότητα ισχύει εάν $D(p||u) = 0$, δηλαδή εάν και μόνο εάν $p(x) = u(x)$.

Ανισότητα **log sum**

- Ανισότητα **log sum**: Για μη αρνητικούς αριθμούς $a_1, a_2, \ldots, a_n$ και $b_1, b_2, \ldots, b_n$,

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}.$$

Η ισότητα ισχύει εάν και μόνο εάν $\frac{a_i}{b_i} = c$, όπου c σταθερά.

Απόδειξη ανισότητας **log sum**

- Απόδειξη: Έστω ότι $a_i > 0$ και $b_i > 0$ (αποδείξτε ως άσκηση την περίπτωση που δεν υπάρχει i για το οποίο να ισχύει $a_i b_i > 0$). Η συνάρτηση $t \log t$ είναι αυστηρώς κυρτή ($\cup$) ($(t \log t)'' = \frac{1}{t} \log e > 0$ για θετικό t). Από την ανισότητα Jensen,
$$\sum \lambda_i f(t_i) \geq f\left(\sum \lambda_i t_i\right),$$
για $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$. Θέτοντας $\lambda_i = \frac{b_i}{\sum_{j=1}^n b_j}$ και $t_j = \frac{a_i}{b_i}$,

$$\sum \frac{a_i}{\sum b_j} \log \frac{a_i}{b_i} \geq \sum \frac{a_i}{\sum b_j} \log \sum \frac{a_i}{\sum b_j} \Rightarrow \sum a_i \log \frac{a_i}{b_i} \geq \left(\sum a_i\right) \log \frac{\sum a_i}{\sum b_i}.$$

Η $D(p||q)$ είναι κυρτή ($\cup$)

- 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.
- Συνεπώς, για την εντροπία ισχύει $H(\lambda p_1 + (1-\lambda)p_2) \geq \lambda H(p_1) + (1-\lambda)H(p_2)$.
- Για εναλλακτική απόδειξη, χωρίς χρήση ανισότητας log sum δείτε Cover Theorem 2.7.3.

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

- Απόδειξη: $I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p(x)H(Y|X=x)$.
- 1ος όρος: $p(y) = \sum_x p(y|x)p(x)$. Συνεπώς, για δεδομένη $p(y|x)$, η $p(y)$ είναι γραμμική συνάρτηση της $p(x)$. Η $H(Y)$ είναι κοίλη συνάρτηση της $p(y)$ και, επομένως, και της $p(x)$.
- 2ος όρος: Γραμμική συνάρτηση της $p(x)$.
- Επομένως, η $I(X;Y)$ είναι κοίλη ($\cap$) συνάρτηση της $p(x)$ για δεδομένη $p(y|x)$.
- Θυμηθείτε ότι σε διακριτά κανάλια χωρίς μνήμη η χωρητικότητα ισούται με τη μέγιστη τιμή της $I(X;Y)$. Το γεγονός ότι η $I(X;Y)$ είναι κοίλη για δεδομένο κανάλι ($p(y|x)$) σημαίνει ότι εάν βρούμε ένα τοπικό μέγιστο, τότε είναι και ολικό μέγιστο και η αντίστοιχη κατανομή πηγής $p^*(x)$ είναι αυτή η οποία επιτυγχάνει τη χωρητικότητα.

Η $I(X;Y)$ είναι κυρτή ($\cup$) συνάρτηση της $p(y|x)$ για δεδομένη $p(x)$

- Έστω δύο υπό συνθήκη κατανομές μάζας πιθανότητας $p_1(y|x)$ και $p_2(y|x)$. $p_1(x,y) = p(x)p_1(y|x)$ και $p_2(x,y) = p(x)p_2(y|x)$. Επίσης, $p_1(y) = \sum_x p_1(x,y)$ και $p_2(y) = \sum_x p_2(x,y)$. Η περιθώρια κατανομή των $p_1(x,y)$ και $p_2(x,y)$ ως προς x είναι η $p(x)$.
- Έστω, τώρα, η υπό συνθήκη κατανομή που προκύπτει από την "ανάμιξη" των $p_1(y|x)$ και $p_2(y|x)$:

$$p_\lambda(y|x) = \lambda p_1(y|x) + (1-\lambda)p_2(y|x),\ 0 \leq \lambda \leq 1.$$

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 $p(x)$ (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 $I(X;Y) = D(p_\lambda(x,y)||p_\lambda(x)p_\lambda(y)) = D(p_\lambda(x,y)||p(x)p_\lambda(y)) = D(p_\lambda(x,y)||q_\lambda(x,y))$.

- Η $D(p||q)$ είναι κυρτή συνάρτηση του ζεύγους (p,q). Επομένως, και η $I(X;Y)$ είναι κυρτή συνάρτηση της $p(y|x)$.

- Συνεπώς, για δεδομένη κατανομή πηγής, υπάρχει κάποιο κανάλι το οποίο ελαχιστοποιεί την πληροφορία που μπορούμε να μεταδώσουμε στο δέκτη.

Ανισότητα Επεξεργασίας Δεδομένων

- Επανάληψη Βασικών Ποσοτήτων Θεωρίας Πληροφορίας (συνέχεια)
- $\boxed{\text{Ανισότητα Επεξεργασίας Δεδομένων}}$
- Ανισότητα Fano
- Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP)

Ανισότητα Επεξεργασίας Δεδομένων

- Οι X, Y, Z σχηματίζουν αλυσίδα Markov ($X \rightarrow Y \rightarrow Z$) εάν $p(x, y, z) = p(x)p(y|x)p(z|y)$.

- Ισοδύναμα, $X \rightarrow Y \rightarrow Z$ 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).

- $X \rightarrow Y \rightarrow g(Y)$.

- Ανισότητα Επεξεργασίας Δεδομένων (Data Processing Inequality): Εάν $X \rightarrow Y \rightarrow Z$, τότε $I(X; Y) \geq I(X; Z)$.

Ανισότητα Επεξεργασίας Δεδομένων (απόδειξη)

- Από τον κανόνα αλυσίδας για την αμοιβαία πληροφορία, $I(X;Y,Z) = I(X;Z) + I(X;Y|Z) = I(X;Y) + I(X;Z|Y) = I(X;Y)$, 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 δε 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)

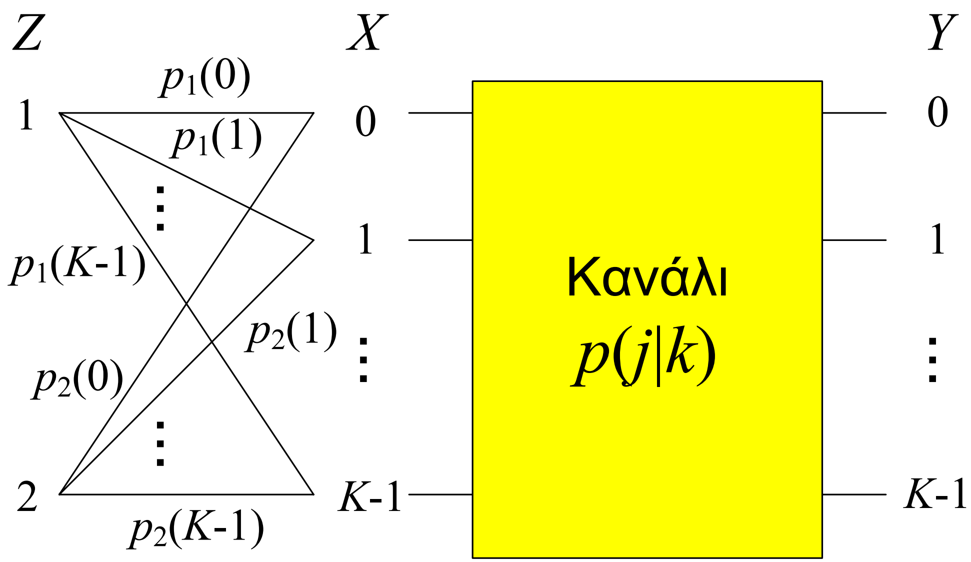
- Μπορούμε να υποθέσουμε ότι οι p_1 και p_2 είναι υπό συνθήκη κατανομές που εξαρτώνται από μια δυαδική τ.μ. Z:

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

- 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,

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

- 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).

Ανισότητα **Fano**

- Επανάληψη Βασικών Ποσοτήτων Θεωρίας Πληροφορίας (συνέχεια)
- Ανισότητα Επεξεργασίας Δεδομένων.
- Ανισότητα Fano
- Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP).

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: $\hat{X} = g(Y)$.

- 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 $X \to Y \to \hat{X}$,

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

  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. σφάλματος

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

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

$$H(E, X|\hat{X}) = H(X|\hat{X}) + \underbrace{H(E|X, \hat{X})}_{=0}$$

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

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

Απόδειξη Ανισότητας **Fano** (συνέχεια)

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

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

Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (**AEP**)

- Επανάληψη Βασικών Ποσοτήτων Θεωρίας Πληροφορίας (συνέχεια)
- Ανισότητα Επεξεργασίας Δεδομένων.
- Ανισότητα Fano
- $\boxed{\text{Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP)}}$

Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (**AEP**) – Εισαγωγή

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

Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (**AEP**) – Εισαγωγή (2)

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

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

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

Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (**AEP**) – Εισαγωγή (3)

- Επομένως, εάν σχηματίσουμε μια ακολουθία πολύ μεγάλου μήκους, η από κοινού συνάρτηση κατανομής μάζας πιθανότητας θα συγκλίνει <u>κατά πιθανότητα</u> στην τιμή $2^{-nH(X)}$.

- Θα αποδείξουμε ότι υπάρχουν περίπου $2^{nH(X)}$ τέτοιες, τυπικές, ακολουθίες και ότι το άθροισμα των από κοινού συναρτήσεων μάζας πιθανότητάς τους προσεγγίζει το 1.

- Το άθροισμα των πιθανοτήτων των υπόλοιπων, μη τυπικών, ακολουθιών μήκους n τείνει στο 0.

- Επομένως, μπορούμε να κωδικοποιήσουμε μόνο τις τυπικές ακολουθίες $\rightarrow$ χρειαζόμαστε nH bits αντί για n.

Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (**AEP**) – Εισαγωγή (4)

- Επειδή η πιθανότητα να εμφανιστεί μη τυπική ακολουθία τείνει στο 0, η πιθανότητα να μη μπορούμε να κωδικοποιήσουμε την ακολουθία X_1^n με χρήση nH bits τείνει στο 0 για $n \to \infty$.

- Το AEP αποτελεί στυλοβάτη της Θεωρίας Πληροφορίας.

Είδη σύγκλισης (υπενθύμιση)

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

1. <u>Κατά πιθανότητα (in probability)</u> εάν, για κάθε $\epsilon > 0$, $\Pr\{|X_n - X| > \epsilon\} \to 0$ για $n \to \infty$.

2. <u>Κατά μέση τετραγωνική τιμή (mean square)</u> εάν $E(X_n - X)^2 \to 0$.

3. <u>Με πιθανότητα 1 (ή σχεδόν βέβαια)</u> εάν $\Pr\{\lim_{n\to\infty} X_n = X\} = 1$.

Τυπικό Σύνολο (**Typical Set**) και ιδιότητες

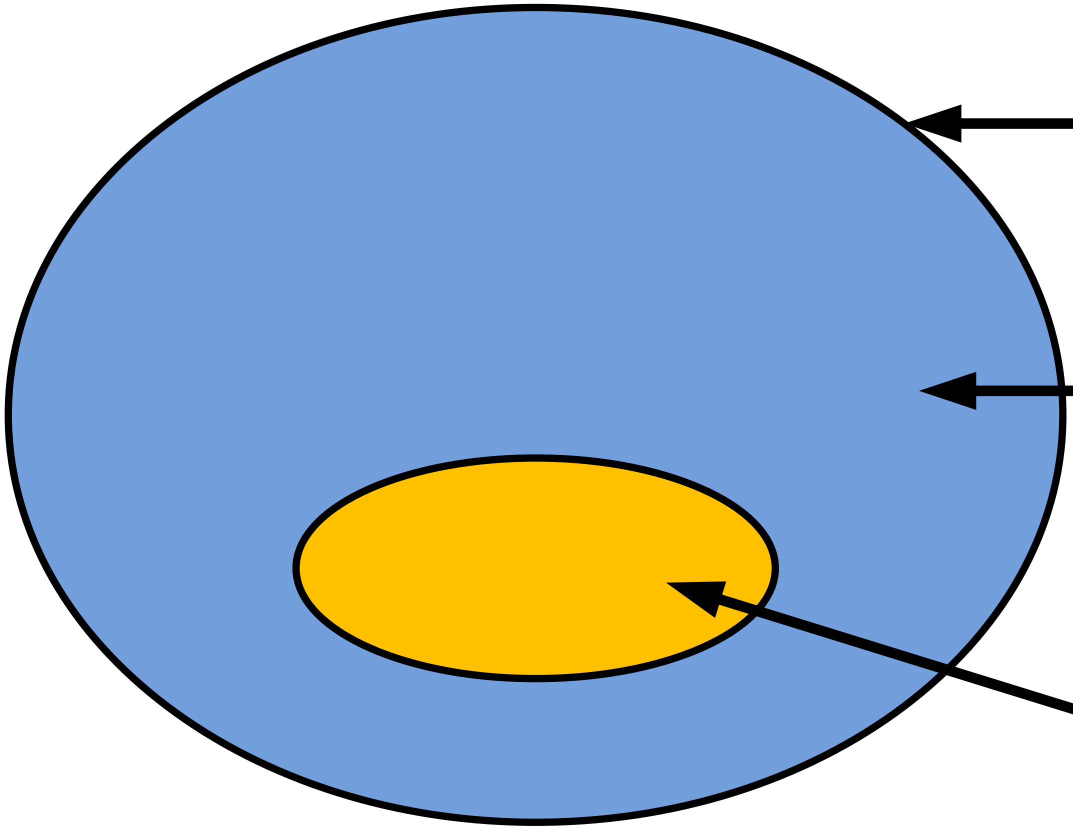
- Το τυπικό σύνολο $A_\epsilon^{(n)}$ που αντιστοιχεί στην κατανομή $p(x)$ αποτελείται από τις ακολουθίες $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ που ικανοποιούν τη σχέση

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

- Ιδιότητες $A_\epsilon^{(n)}$:

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

Τυπικό Σύνολο



$\mathcal{X}^n : |\mathcal{X}|^n$ στοιχεία

Μη τυπικό σύνολο

Τυπικό σύνολο

$A_\epsilon^{(n)} : 2^{n(H+\epsilon)}$ στοιχεία

Ανακεφαλαίωση μαθήματος

- Σχετική Εντροπία και Δεσμευμένη Σχετική Εντροπία· Αμοιβαία Πληροφορία.
- Ανισότητα Πληροφορίας και συνέπειες. Ιδιότητες Εντροπίας και Αμοιβαίας Πληροφορίας.
- Ανισότητα Επεξεργασίας Δεδομένων: Δεν υπάρχει τρόπος επεξεργασίας που να μπορεί να αυξήσει την πληροφορία που περιέχεται σε μια τ.μ. Αντίθετα, ενδέχεται να τη μειώσει.
- Ανισότητα Fano. Δίνει φράγμα για την πιθανότητα σφάλματος στην εκτίμηση τ.μ. με βάση παρατήρηση άλλης τ.μ. Θα τη χρησιμοποιήσουμε εκτενώς στην Κωδικοποίηση Καναλιού.

Ανακεφαλαίωση μαθήματος (συνέχεια)

- Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP): Η "καρδιά" της συμπίεσης (και της Θεωρίας Πληροφορίας). Για μεγάλα μήκη ακολουθιών μπορούμε να χωρίσουμε τις ακολουθίες σε δύο σύνολα
  - Το τυπικό σύνολο. Το άθροισμα των πιθανοτήτων των ακολουθιών που ανήκουν σε αυτό τείνει ασυμπτωτικά στο 1.
  - Το μη τυπικό σύνολο.

Προεπισκόπηση επόμενου μαθήματος

- Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP). Αποδείξεις ιδιοτήτων.
- Εφαρμογή του AEP στην κωδικοποίηση. Κωδικοποίηση σταθερού μήκους.
- Θεώρημα Κωδικοποίησης Πηγής. Απόδειξη για πηγές χωρίς μνήμη.