EE728

Προχωρημένα Θέματα Θεωρίας Πληροφορίας

Δημήτρης - Αλέξανδρος Τουμπακάρης

2ο Μάθημα – 5 Μαρτίου 2008

Ανακεφαλαίωση προηγούμενου μαθήματος

- Πληροφορίες για το μάθημα, τρόπος αξιολόγησης
- Θέματα που θα καλύψουμε στο μάθημα
  - Συμπίεση, AEP και Θεώρημα Κωδικοποίησης Πηγής.
  - Χωρητικότητα Καναλιού, Joint AEP και Θεώρημα Κωδικοποίησης Καναλιού. Κανάλια χωρίς μνήμη με ανάδραση.
  - Συνεχείς τ.μ., Γκαουσιανό Κανάλι.
  - Θεωρία Πληροφορίας Δικτύων (MAC, BC, Relay Channel, Interference Channel).
  - Θεώρημα Κωδικοποίησης Slepian-Wolf.
  - Άλλα θέματα (εάν προλάβουμε).
- Επανάληψη Βασικών Ποσοτήτων Θεωρίας Πληροφορίας
  - Εντροπία (διακριτής τ.μ.)
  - Δεσμευμένη Εντροπία
  - Από κοινού Εντροπία

Προεπισκόπηση σημερινού μαθήματος

- Συνέχεια Επανάληψης και κάποιες αποδείξεις
  - Ρυθμός Εντροπίας
  - Σχετική Εντροπία
  - Αμοιβαία Πληροφορία.
  - Κυρτές συναρτήσεις και ανισότητα Jensen (απόδειξη).
  - Ιδιότητες Εντροπίας, Σχετικής Εντροπίας και Αμοιβαίας Πληροφορίας.

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 $X_i \in \mathcal{X}$. Αποδείχθηκε 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 <u>ανεξάρτητες</u> 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):

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

  εφόσον το όριο συγκλίνει.
- Το όριο συγκλίνει πάντα, εφόσον η πηγή είναι στάσιμη. Στην περίπτωση αυτή, συγκλίνει και η ποσότητα

$$H'(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_n | X_1, X_2, \ldots, X_{n-1})$$

  και $H(\mathcal{X}) = H'(\mathcal{X})$.
- 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 ομοιόμορφα κατανεμημένες, $H(\mathcal{X}) = H'(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} n H(X_i) = H(X_i) = H(X_1)$.
- 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 2.1 (**Cover** σελ. 74)

- Έστω ακολουθία δυαδικών τ.μ. όπου η $p_i = \Pr\{X_i = 1\}$ 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 εκθετικά 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 <u>ίδιο</u> αλφάβητο 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.
- $H(X) + D(p||q) \leq E[l^*] < H(X) + D(p||q) + 1$, 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 μεταβάλλεται σε $p(X|Y)$. Επομένως, κατά μέσο όρο, γνώση της Y ελαττώνει την αβεβαιότητα που έχουμε για την X κατά $E_p\left[\frac{p(X|Y)}{p(X)}\right]$, όπου η μέση τιμή υπολογίζεται για όλες τις τιμές των X και Y.
- Συνεπώς,

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

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

- Προφανώς (από την παραπάνω έκφραση), $I(X;Y) = I(Y;X)$. Άρα, αποκάλυψη της X οδηγεί στην ίδια ελάττωση της αβεβαιότητας για την Y κατά μέσο όρο.
- Η ποσότητα $I(X;Y)$ ονομάζεται αμοιβαία πληροφορία.

Αμοιβαία Πληροφορία $I(X;Y)$ (συνέχεια)

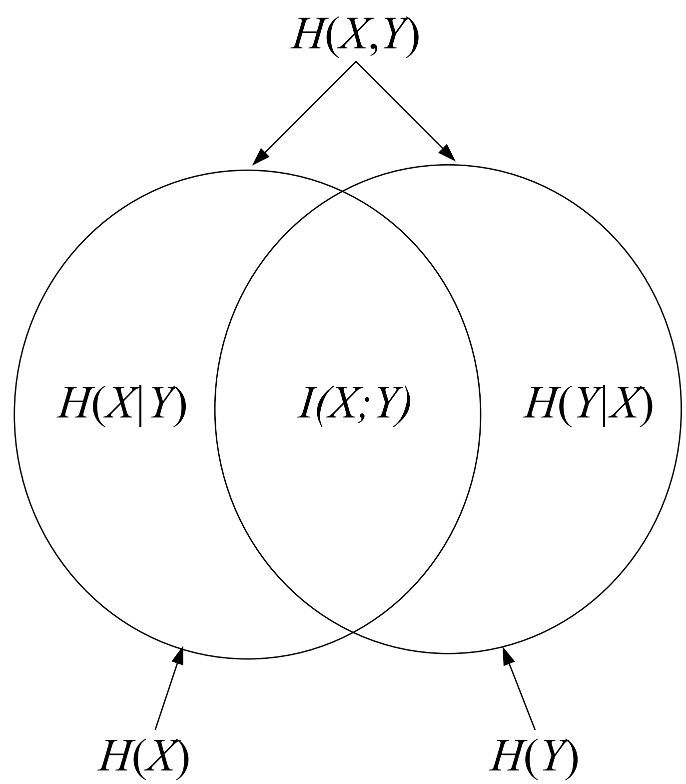
- Μια διαφορετική ερμηνεία της αμοιβαίας πληροφορίας με βάση τη σχετική εντροπία: Η πληροφορία που "χάνουμε" εάν θεωρήσουμε ότι οι X και Y είναι ανεξάρτητες, ενώ, στην πραγματικότητα, δεν είναι.
- $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$. Προκύπτει από τον ορισμό (αποδείχθηκε στη "Θεωρία Πληροφορίας")
- $I(X;X) = H(X) - H(X|X) = H(X)$. 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.



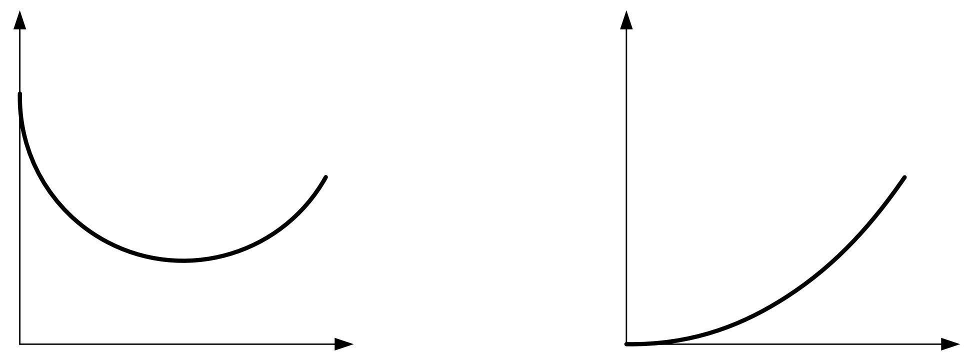
Κυρτές (**convex**) και κοίλες (**concave**) συναρτήσεις

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

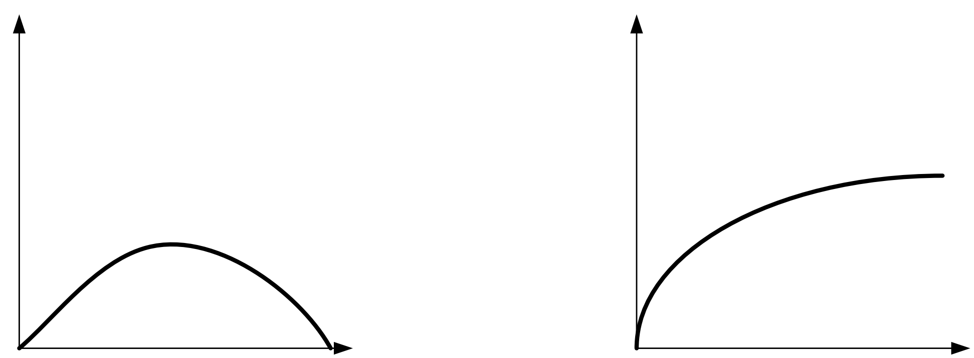
$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

- 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 $x \geq 0$).
- 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 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 για διακριτές τ.μ. (Cover):
  - 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.
    (συνεχίζεται στην επόμενη διαφάνεια)

Συνέχεια απόδειξης ανισότητας **Jensen**

– 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 $A = \{x : p(x) > 0\}$.

$$\begin{aligned} -D(p||q) &= -\sum_{x \in A} p(x) \log \frac{p(x)}{q(x)} = \sum_{x \in A} p(x) \log \frac{q(x)}{p(x)} = \\ &\stackrel{(a)}{\leq} \log \sum_{x \in A} p(x) \frac{q(x)}{p(x)} = \log \sum_{x \in A} q(x) = \\ &\leq \log \sum_{x \in \mathcal{X}} q(x) = \log 1 = 0. \end{aligned}$$

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

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 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 <u>ομοιόμορφα</u> 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: 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.

Προεπισκόπηση επόμενου μαθήματος

- 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 (και απόδειξη).
- Τυπικές ακολουθίες και Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP).
- Εφαρμογή του AEP στην κωδικοποίηση. Κωδικοποίηση σταθερού μήκους.
- Θεώρημα Κωδικοποίησης Πηγής. Δε μπορούμε να συμπιέσουμε περισσότερο από την εντροπία (ή το ρυθμό εντροπίας για εργοδικές πηγές με μνήμη). Απόδειξη για πηγές χωρίς μνήμη.
- Εισαγωγή στην Κωδικοποίηση Καναλιού
  - Διακριτά Κανάλια. Διακριτά Κανάλια Χωρίς Μνήμη.
  - Χωρητικότητα Διακριτού Καναλιού Χωρίς Μνήμη.
  - Συμμετρικά Κανάλια και Χωρητικότητα.