

EE728

Προχωρημένα Θέματα
Θεωρίας Παρηγορίας

Δημήτρης - Αλέξανδρος Τουμπακάκης
3ο Μάθημα – 18 Μαρτίου 2008

Ανακεφαλαίωση προηγούμενου μαθήματος

- Σχετική Εντροπία και Δεσμευμένη Σχετική Εντροπία. Το “κόστος” όταν δε γνωρίζουμε επακριβώς μια κατανομή.
- Αμοιβαία Πληροφορία. Η πληροφορία για μια μεταβλητή X που περιέχεται σε μια μεταβλητή Y (και αντίστροφα).
- Ανισότητα Πληροφορίας και συνέπειες. Η αμοιβαία πληροφορία είναι μη αρνητική. Η δεσμευση ελαττώνει την εντροπία.
- 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 σημερινού μαθήματος

- Η εντροπία είναι κοίλη (\cap) συνάρτηση της $p(x)$.
- Η αμοιβαία πληροφορία είναι κοίλη (\cap) συνάρτηση της $p(x)$ για δεδομένη $p(y|x)$.
- Ανισότητα Επεξεργασίας Δεδομένων
- Ανισότητα Fano
- Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (Asymptotic Equipartition Property).
- Εφαρμογή του AEP στην κωδικοποίηση. Κωδικοποίηση σταθερού μήκους.

Περιεχόμενα σημερινού μαθήματος

- **Ιδιότητες $H(X)$ και $I(X; Y)$ (τέλος!)**
- Ανισότητα επεξεργασίας δεδομένων.
- Ανισότητα Fano.
- Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης (AEP).

Η εντροπία είναι κοίλη (\cap) συνάρτηση της p

- 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 $p(x)$. (Αποδεικνύεται ότι, αν μια συνάρτηση $f(\mathbf{x})$ είναι κυρτή, τότε και η $f(\mathbf{Ax} + \mathbf{b})$ είναι κυρτή).
- 2ος όρος: Γραμμική συνάρτηση της $p(x) \Rightarrow$ κοίλη (και κυρτή).
- Επομένως, η $I(X; Y)$ είναι κοίλη (\cap) συνάρτηση της $p(x)$ για δεδομένη $p(y|x)$.

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

- Θυμηθείτε ότι σε διακριτά κανάλια χωρίς μνήμη η χωρητικότητα ισούται με τη μέγιστη τιμή της $I(X; Y)$ (θα το αποδείξουμε αργότερα στο μάθημα).
- Το γεγονός ότι η $I(X; Y)$ είναι κοίλη για δεδομένο κανάλι, $p(y|x)$, σημαίνει ότι, εάν βρούμε ένα τοπικό μέγιστο (**local optimum**), τότε αυτό είναι και ολικό μέγιστο (**global optimum**) και η αντίστοιχη κατανομή πηγής $p^*(x)$ η οποία επιτυγχάνει το μέγιστο είναι η κατανομή 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 $p_2(y|x)$:

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

Συνεπώς, ισχύει, επίσης,

$$\begin{aligned} p_\lambda(x, y) &= p_\lambda(y|x)p(x) = \lambda p_1(y|x)p(x) + (1 - \lambda)p_2(y|x)p(x) \\ &= \lambda p_1(x, y) + (1 - \lambda)p_2(x, y) \text{ και} \\ p_\lambda(y) &= \lambda p_1(y) + (1 - \lambda)p_2(y). \end{aligned}$$

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 $p(y|x)$.
- Επομένως, για δεδομένη κατανομή πηγής, υπάρχει κάποιο κανάλι το οποίο ελαχιστοποιεί την πληροφορία που μπορούμε να μεταδώσουμε στο δέκτη.

Ανισότητα επεξεργασίας δεδομένων

- Ιδιότητες $H(X)$ και $I(X; Y)$ (ολοκλήρωση)
- 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).
- $X \rightarrow Y \rightarrow g(Y)$ ($g(\cdot)$ οποιαδήποτε νομοτελειωκή (deterministic) συνάρτηση).
- Ανισότητα Επεξεργασίας Δεδομένων (Data Processing Inequality): Εάν $X \rightarrow Y \rightarrow Z$, τότε $I(X; Y) \geq I(X; Z)$.

Ανισότητα Επεξεργασίας Δεδομένων (Απόδειξη)

- 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 δεν μπορεί να αυξηθεί με επεξεργασία της Y (αντίθετα, μάλλον, ενδέχεται να μειωθεί). Ωστόσο, κατάλληλη επεξεργασία της Y ενδέχεται να διευκολύνει την εξαγωγή της πληροφορίας.

Η $I(X; Y)$ είναι κοίλη (Π) συνάρτηση της $p(x)$ για δεδομένη $p(y|x) - \text{Εναλλακτική Απόδειξη (Gallager)}$

- 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 αμοιβαία πληροφορία που αντιστοιχεί στην κατανομή εισόδου p . Θα δείξουμε ότι

$$\theta I_1 + (1 - \theta)I_2 \leq I.$$

Η $I(X; Y)$ είναι κοίλη (Π) συνάρτηση της $p(x)$ για δεδομένη $p(y|x) - \text{Εναλλακτική Απόδειξη (2)}$

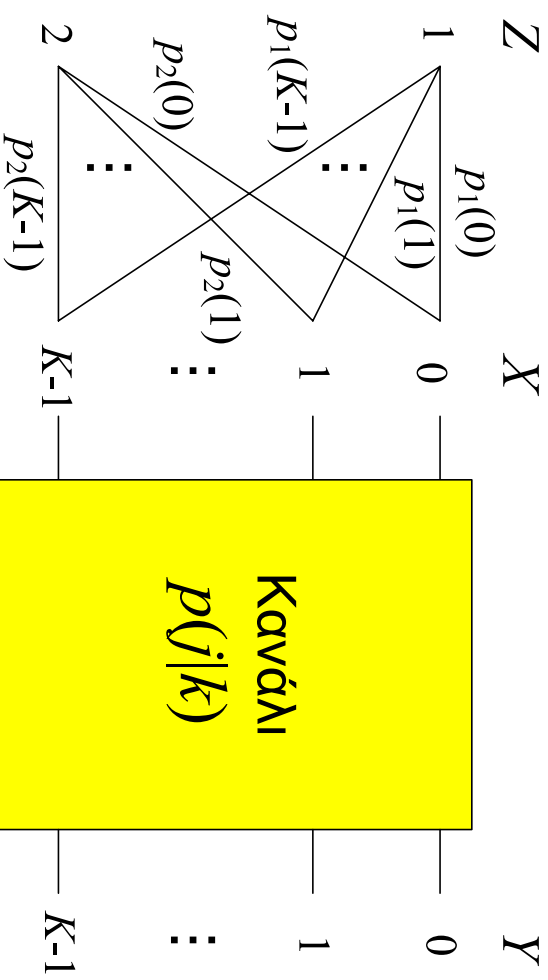
- Μπορούμε να υποθέσουμε ότι οι p_1 και p_2 είναι υπό συνθήκη κατανομές που εξαρτώνται από μια δυαδική τ.μ. Z :

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

- Θέτουμε $p_Z(1) = \theta$ και $p_Z(2) = 1 - \theta$.

Η $I(X; Y)$ είναι κοίλη (Π) συνάρτηση της $p(x)$ για δεδομένη $p(y|x) -$ Εναλλακτική Απόδειξη (3)

Το πρόβλημα φαίνεται στο παρακάτω σχήμα.



Παρατηρούμε ότι $Z \rightarrow X \rightarrow Y$ 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 (Π) συνάρτηση της $p(x)$ για δεδομένη $p(y|x) - \text{Εναλλακτική Απόδειξη (4)}$

- 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 (∪) συνάρτηση της $p(y|x)$ για δεδομένη $p(x)$ (Gallager Theorem 4.4.3).

Ανισότητα Fano

- Ιδιότητες $H(X)$ και $I(X; Y)$ (ολοκλήρωση)
- Ανισότητα επεξεργασίας δεδομένων
- Ανισότητα Fano
- Ιδιότητα Ασυμπτωτικής Ισοδιαμέτρησης (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 (οτότε, 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 $X \rightarrow Y \rightarrow \hat{X}$,

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

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, $P_e = 0 \Rightarrow H(X|Y) = 0$.
- Θέτοντας $H(P_e) = \max_p H(p) = 1$ προκύπτει ένα 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. που δηλώνει εάν έχει συμβεί σφάλμα εκτίμησης

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

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

$$\begin{aligned} H(E, X|\hat{X}) &= H(X|\hat{X}) + \underbrace{H(E|X, \hat{X})}_{=0} \\ &= \underbrace{H(E|\hat{X})}_{\leq H(E)=H(P_e)} + \underbrace{H(X|E, \hat{X})}_{\leq P_e \log |\mathcal{X}|}. \end{aligned}$$

Απόδειξη Ανισότητας Fano (2)

$$H(E, X|\hat{X}) = H(X|\hat{X}) + \underbrace{H(E|X, \hat{X})}_{=0} = \underbrace{H(E|\hat{X})}_{\leq H(E)=H(P_e)} + \underbrace{H(X|E, \hat{X})}_{\leq P_e \log |\mathcal{X}|}.$$

- $H(E|X, \hat{X}) = 0$ γιατί εάν ξέρουμε τις τιμές των \hat{X} και X γνωρίζουμε εάν έχει εμφανιστεί σφάλμα εκτίμησης.
- $H(E|\hat{X}) \leq H(E)$. Δεδομένου ότι η πιθανότητα σφάλματος ($E = 1$) ισούται με P_e , η τ.μ. ακολουθεί κατανομή Bernoulli με ταχύμετρο 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 $H(X|E, \hat{X}) = H(X)$ και η X να ακολουθεί ομοιόμορφη κατανομή.

Απόδειξη Ανισότητας Fano (3)

- Επομένως, $H(P_e) + P_e \log |\mathcal{X}| \geq H(X|\hat{X})$.
- Δεδομένου ότι $X \rightarrow Y \rightarrow \hat{X}$, $I(X;Y) \geq I(X;\hat{X}) \Rightarrow H(X) - H(X|Y) \geq H(X) - H(X|\hat{X}) \Rightarrow H(X|\hat{X}) \geq H(X|Y)$. 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} ,
$$H(X|E, \hat{X}) \leq P_e \log(|\mathcal{X}| - 1) \text{ και}$$

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

Ιδιότητα Ασυμπτωτικής Ισοδιαμέτρησης (**AEP**)

- Ιδιότητες $H(X)$ και $I(X; Y)$ (ολοκλήρωση)
- Ανισότητα επεξεργασίας δεδομένων
- Ανισότητα Fano
- Ιδιότητα Ασυμπτωτικής Ισοδιαμέτρησης (**AEP**)

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

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

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

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

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

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

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

- Επομένως, εάν σχηματίσουμε μια ακολουθία πολύ μεγάλου μήκους, η από κοινού συνάρτηση κατανομής μάζας πιθανότητας θα συγκλίνει κατά πιθανότητα 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.

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

- Επομένως, μπορούμε να κωδικοποιήσουμε μόνο τις τυπικές ακολουθίες \rightarrow χρειαζόμαστε nH bits αντί για n .
- 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 $n \rightarrow \infty$.
- Το **AEP** αποτελεί έναν από τους στυλοβάτες της Θεωρίας Πληροφορίας.

Προπαισκότηση επόμενου μαθήματος

- Ιδιότητα Ασυμπτωτικής Ισοδιαμέτρησης (AEP) (συνέχεια). Ορισμός τυπικών ακολουθιών και ιδιότητες.
- Θέωρημα Κωδικοποίησης Πηγής. Απόδειξη για πηγές χωρίς μνήμη.
- Εισαγωγή στην Κωδικοποίηση Καναλιού
 - Διακριτά Κανάλια. Διακριτά Κανάλια Χωρίς Μνήμη.
 - Πληροφοριακή Χωρητικότητα Διακριτού Καναλιού Χωρίς Μνήμη.
 - Συμμετρικά Κανάλια και Χωρητικότητα.
- Από Κοινού Τυπικότητα και Ιδιότητα Από Κοινού Ασυμπτωτικής Ισοδιαμέτρησης (Joint AEP).
- Εισαγωγή στο Θέωρημα Κωδικοποίησης Καναλιού και 1ο μέρος απόδειξης.