EE728
Προχωρημένα Θέματα Θεωρίας Πληροφορίας
5η διάλεξη
2η έκδοση, 27/4/2015

Δημήτρης-Αλέξανδρος Τουμπακάρης

Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών

21 Απριλίου 2015

Περιεχόμενα 5ης διάλεξης

Αντιστοιχία 5ης διάλεξης με βιβλία Cover & Thomas και El Gamal & Kim

- Βιβλίο Cover & Thomas (2η έκδοση): 2.8, 2.10, 7.1–7.5.
- Βιβλίο El Gamal & Kim: 3.1–3.1.1, 3.5.

Η Ανισότητα Επεξεργασίας Δεδομένων και η 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).
  - Ισχύει πάντοτε ότι $X \to Y \to g(Y)$.
- Ανισότητα Επεξεργασίας Δεδομένων (Data Processing Inequality):

Ανισότητα Επεξεργασίας Δεδομένων

Εάν $X \to Y \to Z$, τότε $I(X; Y) \geq I(X; Z)$.

Ανισότητα Επεξεργασίας Δεδομένων (απόδειξη)

- 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 της 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 η αμοιβαία πληροφορία μεταξύ 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)

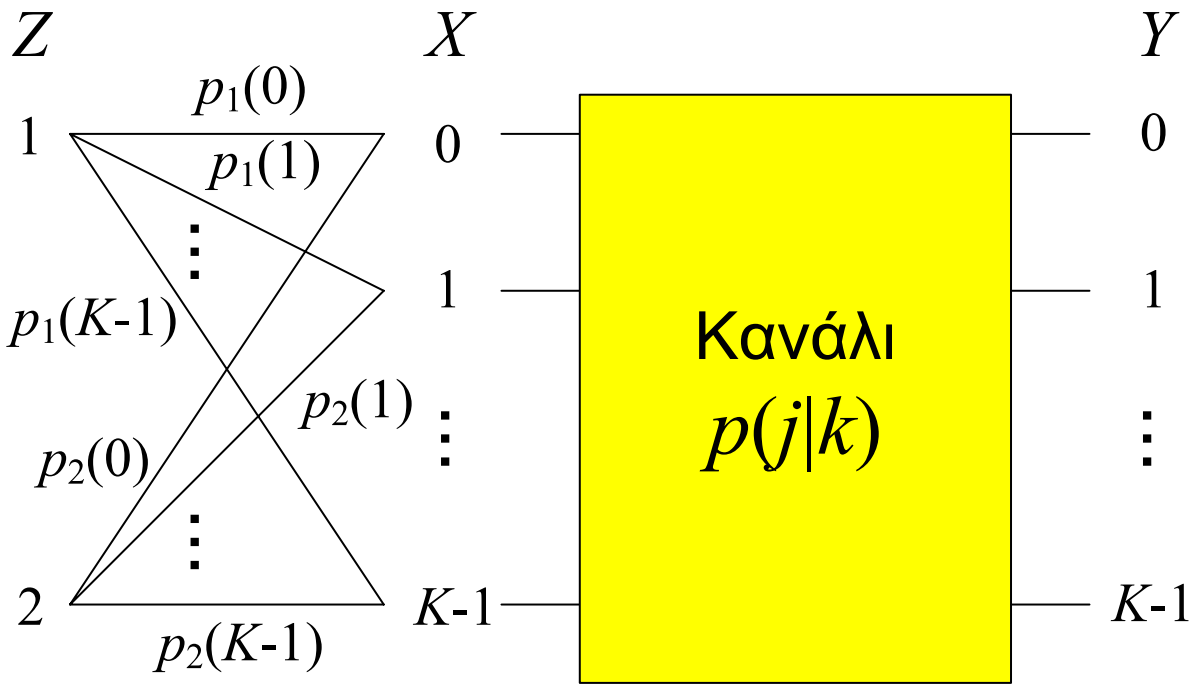
- Μπορούμε να υποθέσουμε ότι οι p_1 και p_2 είναι υπό συνθήκη κατανομές που εξαρτώνται από μια δυαδική τ.μ. Z:

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

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

Η Ανισότητα Fano

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 ντετερμινιστικός (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 $P_e \triangleq \Pr\{\hat{X} \neq X\}$.

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, ντετερμινιστική 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 και στην απόδειξη του Θεωρήματος Κωδικοποίησης Πηγής (αντίστροφο).
- Γενικώς, η ανισότητα Fano χρησιμοποιείται στις περισσότερες αποδείξεις αντιστρόφων (converse proofs).

Απόδειξη Ανισότητας Fano

(Cover & Thomas 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)

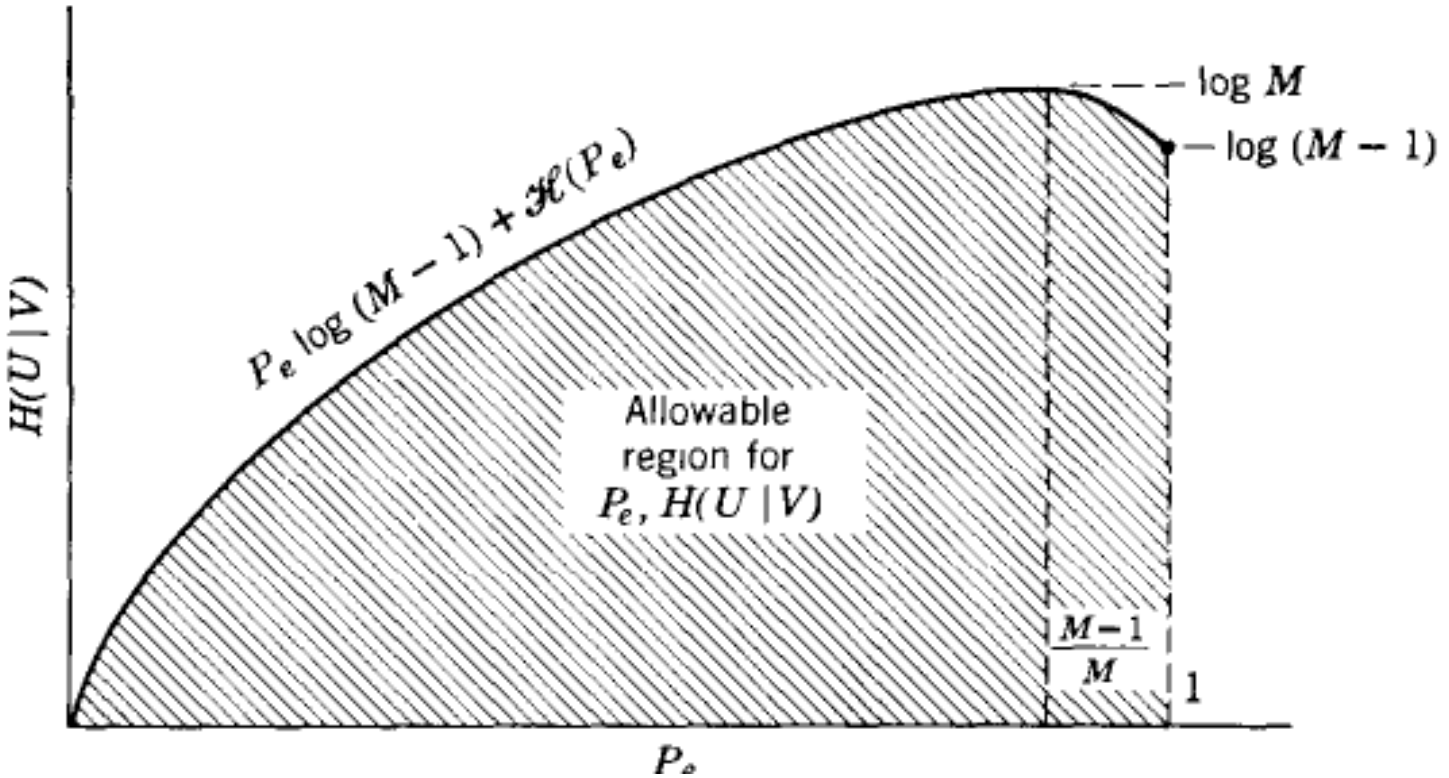
- $H(P_e) + P_e \log |\mathcal{X}| \geq H(X|\hat{X})$.
- Δεδομένου ότι $X \to Y \to \hat{X}$,
  $I(X;Y) \geq I(X;\hat{X}) \Rightarrow H(X) - H(X|Y) \geq H(X) - H(X|\hat{X}) \Rightarrow$
  $H(X|\hat{X}) \geq H(X|Y)$.
  Συνεπώς,

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

- 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 $P_e, H(X|Y)$



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

Κωδικοποίηση Σταθερού Μήκους και Θεώρημα Κωδικοποίησης Πηγής

Κωδικοποίηση Σταθερού Μήκους

- Έστω, 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 $A_\epsilon^{(n)^c} = \mathcal{X}^n - A_\epsilon^{(n)}$.
- Κατασκευή βιβλίου κωδίκων (codebook): Διατάσσουμε όλες τις τυπικές ακολουθίες (π.χ. με αλφαβητική σειρά) και σε κάθε ακολουθία αντιστοιχίζουμε μία κωδική λέξη μήκους L.
- Δεδομένου ότι το τυπικό σύνολο περιέχει το πολύ $2^{n(H+\epsilon)}$ ακολουθίες (σύμφωνα με την Ιδιότητα 3), χρειαζόμαστε το πολύ $L = n(H+\epsilon) + 1$ bits για να τις αναπαραστήσουμε (το επιπλέον 1 bit οφείλεται στο ότι ενδέχεται η ποσότητα $n(H+\epsilon)$ να μην είναι ακέραιος).
- Όλες οι κωδικές λέξεις έχουν το *ίδιο* μήκος L.

Κωδικοποίηση Σταθερού Μήκους (2)

- Σχηματίζουμε ακολουθία μήκους $n > n_0$ από τα σύμβολα X_i της πηγής που θέλουμε να κωδικοποιήσουμε.
- Κωδικοποίηση (encoding):
  - Εάν η ακολουθία είναι τυπική, την κωδικοποιούμε με την κωδική λέξη μήκους L του βιβλίου κωδίκων.
  - Εάν η ακολουθία δεν είναι τυπική, η κωδικοποίηση αποτυγχάνει.
  - Μπορούμε να ελαττώσουμε την πιθανότητα αποτυχίας, ϵ, όσο θέλουμε αυξάνοντας το μήκος, n, των ακολουθιών που κωδικοποιούμε.
- Επομένως, μπορούμε να κωδικοποιήσουμε με χρήση $L/n = (H + \epsilon) + 1/n$ bits/σύμβολο πηγής και να διασφαλίσουμε ότι η πιθανότητα αποτυχίας είναι μικρότερη του ϵ.

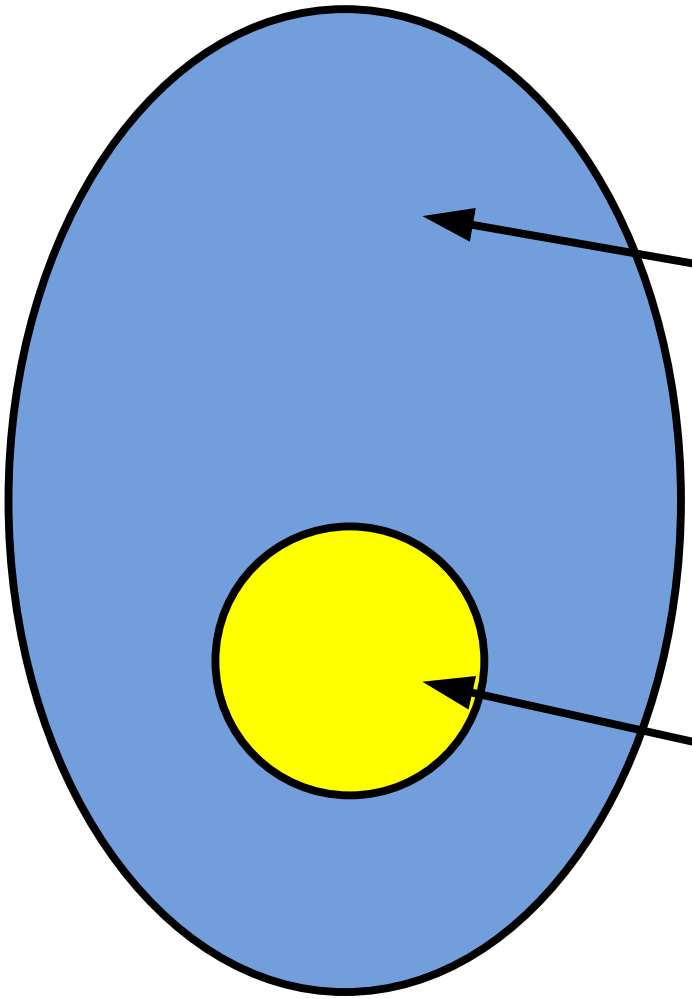
Κωδικοποίηση Σταθερού Μήκους (3)

- Με μία μικρή αλλαγή στον τρόπο κωδικοποίησης μπορούμε να διασφαλίσουμε ότι η πιθανότητα αποτυχίας κωδικοποίησης είναι ακριβώς ίση με 0.
- Ωστόσο, στην περίπτωση αυτή, η κωδικοποίηση δεν είναι σταθερού μήκους.
- Παρατηρούμε ότι για να περιγράψουμε τις ακολουθίες του μη τυπικού συνόλου χρειαζόμαστε το πολύ $n \log |\mathcal{X}| + 1$ bits.
- Διατηρούμε το βιβλίο κωδίκων των τυπικών ακολουθιών και προσθέτουμε και ένα βιβλίο κωδίκων για τις μη τυπικές ακολουθίες.
- Το βιβλίο κωδίκων για τις μη τυπικές ακολουθίες μπορεί να είναι τετριμμένο, δηλαδή να μη συμπιέζουμε την ακολουθία.

Κωδικοποίηση Σταθερού Μήκους (4)

- Κωδικοποίηση:
  - 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 και το βιβλίο κωδίκων των τυπικών ακολουθιών (μήκους L). Επομένως, χρειαζόμαστε $L + 1 = n(H(X) + \epsilon) + 2$ bits.
  - Αλλιώς, αν η ακολουθία είναι μη τυπική, χρησιμοποιούμε πρόθεμα 1 και, στη, συνέχεια, την ίδια την ακολουθία (χωρίς να τη συμπιέσουμε). Επομένως, χρειαζόμαστε $n \log |\mathcal{X}| + 2$ bits.

Κωδικοποίηση Σταθερού Μήκους με χρήση τυπικού συνόλου



Μη τυπικό σύνολο

Περιγραφή: $n\log|\mathcal{X}|+2$ bits

Τυπικό σύνολο

Περιγραφή: $n(H+\varepsilon)+2$ bits

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 $\sim nH$ 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 πιο αποδοτική (χρησιμοποιώντας, π.χ. $n \log \left| A_\epsilon^{(n)^c} \right|$ bits).
- Παρατηρήστε ότι στο όριο το τυπικό σύνολο περιέχει πολύ μικρό ποσοστό των ακολουθιών (το μέγεθός του είναι $\sim 2^{nH}$ οπότε περιέχει ποσοστό $\sim 2^{n(H(X)-\log|\mathcal{X}|)} \to 0$ για $n \to \infty$ εάν $H(X) < \log|\mathcal{X}| - \epsilon$). 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

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, $H(X) \leq \mathbb{E}[l^*] < H(X)+1 \Rightarrow H(X^n) \leq \mathbb{E}[\tilde{l}^*] < H(X^n)+1 \Rightarrow nH(X) \leq \mathbb{E}[\tilde{l}^*] < nH(X)+1 \Rightarrow H(X) \leq \mathbb{E}[\tilde{l}^*]/n < H(X)+1/n$.
- Επομένως, υπάρχει και δεύτερος τρόπος να συμπιέσουμε κοντά στην εντροπία, αυτή τη φορά με κώδικα μεταβλητού μήκους.

Θεώρημα Κωδικοποίησης Πηγής (2)

- Οι δύο τρόποι κωδικοποίησης που προαναφέρθηκαν αποτελούν αποδείξεις της *επιτευξιμότητας* (achievability) του Θεωρήματος Κωδικοποίησης Πηγής για πηγές χωρίς μνήμη (το οποίο, επίσης, ονομάζεται ευθύ μέρος του Θεωρήματος).
- Ωστόσο, για να αποδειχθεί το Θεώρημα Κωδικοποίησης Πηγής πρέπει, επίσης, να δείξουμε ότι δεν υπάρχει τρόπος να συμπιέσουμε περισσότερο τα σύμβολα της πηγής (αντίστροφο (converse) του Θεωρήματος).
- Ένας άλλος τρόπος να το σκεφτούμε είναι ο εξής: Η επιτευξιμότητα μας δίνει ένα άνω φράγμα για το μήκος της περιγραφής της συμπιεσμένης ακολουθίας. Αν βρούμε ένα κάτω φράγμα το οποίο ταυτίζεται με το άνω φράγμα έχουμε αποδείξει το Θεώρημα.
- Θα αποδείξουμε ότι, εάν προσπαθήσουμε να συμπιέσουμε με μέσο μήκος μικρότερο από την εντροπία, η πιθανότητα αδυναμίας αποκωδικοποίησης $P_e \to 1$.

Θεώρημα Κωδικοποίησης Πηγής (3) – αντίστροφο

- Δείτε π.χ. R. Gallager, *Information Theory and Reliable Communication*, R. W. Yeung, *A First Course in Information Theory*.
- Έστω ότι το μήκος της αρχικής (προς συμπίεση) ακολουθίας ισούται με n. Θεωρούμε δυαδικές ακολουθίες (αν και η απόδειξη γενικεύεται εύκολα). Έστω ότι η ακολουθία συμπιέζεται με χρήση L bits, όπου $L < n[H(X) - \zeta]$, $\zeta > 0$ και ότι το ζ δε μεταβάλλεται με το n. Επομένως, μπορούμε να ανακατασκευάσουμε το πολύ $M = 2^{n(H(X)-\zeta)}$ ακολουθίες στην έξοδο του αποκωδικοποιητή.
- Έστω ότι αντιστοιχίζουμε κάποιες από τις M κωδικές λέξεις σε τυπικές ακολουθίες και κάποιες σε μη τυπικές.
- Το άθροισμα των μαζών πιθανότητας των τυπικών ακολουθιών που μπορούμε να κωδικοποιήσουμε δεν μπορεί να υπερβαίνει την τιμή

$$2^{n[H(X)-\zeta]}2^{-n[H(x)-\epsilon]} = 2^{-n[\zeta-\epsilon]}.$$

Θεώρημα Κωδικοποίησης Πηγής (4) – αντίστροφο

- Επομένως, το άθροισμα των μαζών πιθανότητας όλων των M ακολουθιών που μπορούμε να κωδικοποιήσουμε δεν μπορεί να υπερβαίνει την τιμή

$$\begin{aligned} & 2^{n[H(X)-\zeta]}2^{-n[H(x)-\epsilon]} + \Pr\{X_1^n \notin A_\epsilon^{(n)}\} \\ & = 2^{-n[\zeta-\epsilon]} + \Pr\{X_1^n \notin A_\epsilon^{(n)}\} \\ & \overset{(a)}{<} 2^{-n[\zeta-\epsilon]} + \epsilon. \end{aligned}$$

  (a) από το AEP, για αρκούντως μεγάλο n.

- Συνεπώς, για την πιθανότητα να μην έχουμε κατασκευάσει (δηλαδή να μην υπάρχει διαθέσιμη) κωδική λέξη για μία ακολουθία ισχύει

$$P_e^{(n)} > 1 - 2^{-n[\zeta-\epsilon]} - \epsilon,$$

  για αρκούντως μεγάλο n.

Θεώρημα Κωδικοποίησης Πηγής (5) – αντίστροφο

$$P_e^{(n)} > 1 - 2^{-n[\zeta-\epsilon]} - \epsilon,$$

για αρκούντως μεγάλο n, για οποιοδήποτε $\epsilon > 0$.

- Άρα, ισχύει και για $\epsilon < \zeta$.
- Αλλά για οποιοδήποτε $\epsilon < \zeta$, $P_e > 1 - 2\epsilon$ για αρκούντως μεγάλο n.
- Συνεπώς, $P_e^{(n)} \to 1$ για $n \to \infty$, αφού, για μεγάλο n, και $\epsilon \to 0$.

Θεώρημα Κωδικοποίησης Πηγής (6) – αντίστροφο

- Ένα ερώτημα που προκύπτει εδώ είναι το εξής: Δείξαμε ότι $P_e^{(n)} \to 1$ για $n \to \infty$. Θα μπορούσε κάποιος να ισχυριστεί ότι ίσως να υπάρχει κάποιος τρόπος να κωδικοποιήσουμε με κάποια πεπερασμένη τιμή n και με τον τρόπο αυτό να επιτύχουμε συμπίεση με μέσο μήκος μικρότερο από την εντροπία.
- Μπορούμε να δείξουμε ότι κάτι τέτοιο δεν είναι δυνατό.
  - Έστω ότι υπάρχει τρόπος κωδικοποίησης με κάποιο (σχετικά μικρό) n για τον οποίο $P_e^{(n)} \to 0$.
  - Έστω, τώρα, ότι θέλουμε να κωδικοποιήσουμε μία ακολουθία μήκους Kn, $K \to \infty$. Ένας τρόπος να το επιτύχουμε είναι χωρίζοντάς την σε ακολουθίες μήκους n και χρησιμοποιώντας τη μέθοδο που επιτυγχάνει $P_e^{(n)} \to 0$.
  - Ωστόσο, αυτό σημαίνει ότι βρήκαμε έναν τρόπο να κατασκευάσουμε κώδικα μήκους Kn για τον οποίο $P_e^{(Kn)} \to 0$.
  - Αλλά αυτό είναι άτοπο γιατί δείξαμε ότι, για $n \to \infty$, $P_e^{(n)} \to 1$.

Ασθενές Αντίστροφο Θεώρηματος Κωδικοποίησης Πηγής

- Παρατηρήστε ότι αποδείξαμε όχι μόνο ότι δεν μπορούμε να συμπιέσουμε με ρυθμό μικρότερο από την εντροπία, αλλά και ότι, αν προσπαθήσουμε να συμπιέσουμε με $H(X) - \zeta$, $\zeta > 0$, η πιθανότητα αποτυχίας αποκωδικοποίησης τείνει στο 1.
- Αυτό ονομάζεται *ισχυρό αντίστροφο* (strong converse).
- Θα αποδείξουμε, επίσης, το ασθενές αντίστροφο (weak converse) ότι, δηλαδή, δεν υπάρχει κώδικας με μέσο μήκος μικρότερο από την εντροπία ο οποίος να επιτυγχάνει αυθαίρετα μικρή πιθανότητα αποτυχίας κωδικοποίησης.
- Το ασθενές αντίστροφο προκύπτει από το ισχυρό. Ο λόγος που θα κάνουμε την απόδειξη είναι για να εξοικειωθούμε με τη χρήση της Ανισότητας Fano στην απόδειξη ασθενώς αντιστρόφων.

Ασθενές Αντίστροφο Θεώρηματος Κωδικοποίησης Πηγής (2)

- Έστω ότι κατασκευάζουμε έναν κώδικα συμπίεσης για M ακολουθίες πηγής μήκους n. Επομένως, μπορούμε να γράψουμε $M = 2^{nR}$ όπου R ο μέσος αριθμός των bits ανά σύμβολο πηγής.
- Παρατηρούμε ότι $X_1^n \to M \to \hat{X}_1^n$, όπου X_1^n η ακολουθία που παράγει η πηγή, M ο δείκτης της κωδικής λέξης στο βιβλίο κωδίκων του συμπιεστή (κωδικοποιητή πηγής) και $\hat{X}_1^n$ η αποσυμπιεσμένη ακολουθία στο δέκτη.
- Επομένως, από την Ανισότητα Επεξεργασίας Δεδομένων,

$$
\begin{aligned}
&I(X_1^n; M) \geq I(X_1^n; \hat{X}_1^n) \Rightarrow \\
&H(X_1^n) - H(X_1^n|M) \geq H(X_1^n) - H(X_1^n|\hat{X}_1^n) \Rightarrow \\
&H(X_1^n|M) \leq H(X_1^n|\hat{X}_1^n).
\end{aligned}
$$

Ασθενές Αντίστροφο Θεώρηματος Κωδικοποίησης Πηγής (3)

- Από την Ανισότητα Fano,

$$H(X_1^n|M) \leq H(X_1^n|\hat{X}_1^n) \leq nP_e^{(n)} \log|\mathcal{X}| + 1$$
$$= n\left(P_e^{(n)} \log|\mathcal{X}| + \frac{1}{n}\right) \triangleq n\epsilon_n.$$

- Επειδή θέλουμε ο κώδικας να επιτυγχάνει $P_e^{(n)} \to 0$ για $n \to \infty$, $\epsilon_n \to 0$ για $n \to \infty$.

Ασθενές Αντίστροφο Θεώρηματος Κωδικοποίησης Πηγής (4)

- Επομένως

$$\begin{aligned} nR &\overset{(a)}{\geq} H(M) \overset{(b)}{=} H(M) - H(M|X_1^n) \\ &= I(M; X_1^n) = H(X_1^n) - H(X_1^n|M) \\ &\overset{(c)}{\geq} nH(X) - n\epsilon_n. \end{aligned}$$

  (a) $M = 2^{nR}$. (b) Ο δείκτης, M, της κωδικής λέξης είναι ντετερμινιστική συνάρτηση της ακολουθίας X_1^n της πηγής. (c) Από την Ανισότητα Fano.

- Συνεπώς, για $n \to \infty$, $R \geq H(X)$.
- Η Ανισότητα Fano είναι ιδιαίτερα χρήσιμη στην απόδειξη του ασθενούς αντιστρόφου. Θα την χρησιμοποιήσουμε ξανά στα κανάλια.

Θεώρημα Κωδικοποίησης Πηγής (7)

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

Αποδοτική Κωδικοποίηση Πηγής

- Έστω ότι, με χρήση κώδικα, η ακολουθία $(X_1, X_2, \ldots, X_n)$ μίας πηγής κωδικοποιείται στη *δυαδική* ακολουθία $(Y_1, Y_2, \ldots, Y_m)$. Θεωρούμε ότι οι X_i είναι i.i.d. (όχι, απαραίτητα, δυαδικές), δηλαδή ότι η πηγή δεν έχει μνήμη.
- Έστω, επίσης, ότι το αλφάβητο $\mathcal{X}$ της πηγής είναι πεπερασμένο (για απλοποίηση).
- Από το AEP, για $n \to \infty$, $m \approx nH(X)$.
- Εάν $\hat{X}_1^n$ είναι η ανακατασκευασμένη (αποσυμπιεσμένη) ακολουθία, η πιθανότητα εσφαλμένης αποκωδικοποίησης είναι $P_e = \Pr\{X_1^n \neq \hat{X}_1^n\}$.
- Θα δείξουμε ότι, εάν απαιτήσουμε $P_e \to 0$ για $n \to \infty$, τα σύμβολα Y_i της ακολουθίας Y_1^m είναι (σχεδόν) i.i.d. Bern$(1/2)$.

Αποδοτική Κωδικοποίηση Πηγής (2)

- Από την ανισότητα Fano,

$$H(X_1^n|\hat{X}_1^n) \leq 1 + P_e \log |\mathcal{X}|^n = 1 + nP_e \log |\mathcal{X}|.$$

- Επειδή $\hat{X}_1^n = f(Y_1^m)$, $H(Y_1^m) = H(Y_1^m, \hat{X}_1^n) \geq H(\hat{X}_1^n)$.
- Επομένως,

$$\begin{aligned} H(Y_1^m) &\geq H(\hat{X}_1^n) \geq H(\hat{X}_1^n) - H(\hat{X}_1^n|X_1^n) \\ &= I(X_1^n; \hat{X}_1^n) = H(X_1^n) - H(X_1^n|\hat{X}_1^n) \\ &= nH(X) - H(X_1^n|\hat{X}_1^n) \\ &\overset{(a)}{\geq} nH(X) - (1 + nP_e \log |\mathcal{X}|) \\ &= n(H(X) - P_e \log |\mathcal{X}|) - 1. \end{aligned}$$

(a) Ανισότητα Fano.

Αποδοτική Κωδικοποίηση Πηγής (3)

$$H(Y_1^m) \geq n(H(X) - P_e \log |\mathcal{X}|) - 1.$$

- Επίσης, από το φράγμα ανεξαρτησίας της εντροπίας,

$$H(Y_1^m) \leq \sum_{i=1}^{m} H(Y_i) \leq m,$$

  επειδή έχουμε υποθέσει ότι οι Y_i είναι δυαδικές.
- Συνεπώς,

$$n(H(X) - P_e \log |\mathcal{X}|) - 1 \leq H(Y_1^m) \leq m.$$

- Αλλά για $P_e \to 0$ και $n \to \infty$, το κάτω φράγμα τείνει στο $nH(X) \approx m$.

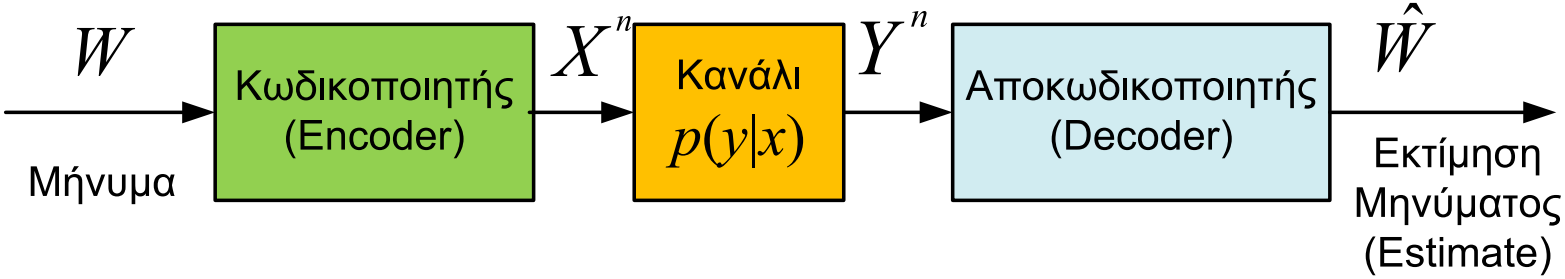
Αποδοτική Κωδικοποίηση Πηγής (4)

- Επομένως, $H(Y_1^m) \approx m$.
- Δηλαδή, η ακολουθία Y_1^m έχει τη μέγιστη δυνατή εντροπία (είναι όσο πιο τυχαία γίνεται).
- Διαισθητικά, αν η $H(Y_1^m)$ δεν ήταν εντελώς τυχαία, θα μπορούσαμε να τη συμπιέσουμε περισσότερο, οπότε ο τρόπος που χρησιμοποιήσαμε αρχικά δε θα ήταν βέλτιστος.

Διακριτά κανάλια και χωρητικότητα

Διακριτά Κανάλια – Εισαγωγή

- Έως τώρα το ενδιαφέρον εστιάστηκε στη βέλτιστη συμπίεση της πληροφορίας που παράγει μια πηγή.
- Το δεύτερο μεγάλο κεφάλαιο της Θεωρίας Πληροφορίας ασχολείται με τη μετάδοση πληροφορίας μέσω ενός καναλιού.



- Στο σχήμα, η πηγή θέλει να μεταδώσει ένα μήνυμα W μέσω ενός καναλιού. Το κανάλι παραμορφώνει/αλλάζει το μήνυμα.

Διακριτά Κανάλια – Εισαγωγή (2)

- Ποιος είναι ο μέγιστος αριθμός διαφορετικών *ανεξάρτητων* μηνυμάτων που μπορούμε να μεταδώσουμε στο κανάλι έτσι ώστε να μπορούμε να τα ανακτήσουμε με αυθαίρετα μικρή πιθανότητα σφάλματος στο δέκτη;
- Πώς εξαρτάται ο μέγιστος αριθμός από τα χαρακτηριστικά του καναλιού;
- Πώς επιτυγχάνεται μετάδοση του μέγιστου αυτού αριθμού μηνυμάτων;
- Πόση είναι η μέγιστη πληροφορία που μπορεί να μεταδοθεί στο κανάλι;

Διακριτά Κανάλια – Εισαγωγή (3)

- Στο μάθημα "Θεωρία Πληροφορίας" είδαμε ότι, για κανάλια *χωρίς μνήμη*, ο μέγιστος ρυθμός μετάδοσης για τον οποίο η πιθανότητα σφάλματος στο δέκτη είναι αυθαίρετα μικρή ονομάζεται χωρητικότητα (capacity) του καναλιού C και ισούται (για κανάλι διακριτού χρόνου) με $\max_{p(x)} I(X;Y)$ (Θεώρημα Χωρητικότητας Καναλιού – θα το αποδείξουμε σύντομα).
- Σε κανάλια με μνήμη ο χαρακτηρισμός είναι πιο σύνθετος και δεν ορίζεται πάντοτε μία μοναδική τιμή χωρητικότητας.
  - Δείτε π.χ. Gallager 4.6

Διακριτά Κανάλια – Εισαγωγή (4)

- Εάν θέλουμε να μεταδώσουμε την πληροφορία μιας πηγής μέσα από ένα κανάλι δεν είναι προφανές εάν η συμπίεση της πηγής θα πρέπει να γίνει λαμβάνοντας υπόψη το κανάλι στο οποίο θα μεταδοθεί η πληροφορία ή εάν η κωδικοποίηση πηγής και η κωδικοποίηση καναλιού μπορούν να γίνουν ανεξάρτητα. Το Θεώρημα Διαχωρισμού Πηγής-Καναλιού εξασφαλίζει ότι οι δύο κωδικοποιήσεις μπορούν να γίνουν ανεξάρτητα στην περίπτωση διακριτού καναλιού χωρίς μνήμη.
- Επομένως, εάν για μια πηγή ισχύει $H(\mathcal{X}) < C$, η πληροφορία που παράγει η πηγή μπορεί να μεταδοθεί μέσω του καναλιού με αυθαίρετα μικρή πιθανότητα σφάλματος.
- Το Θεώρημα Διαχωρισμού Πηγής-Καναλιού ενοποιεί τη συμπίεση και την κωδικοποίηση καναλιού (για διακριτά κανάλια ενός χρήστη, χωρίς μνήμη).

Διακριτά κανάλια – Ορισμοί

- Ορισμός 5.1. Ένα διακριτό κανάλι $(\mathcal{X}, p(y|x), \mathcal{Y})$ αποτελείται από δύο πεπερασμένα σύνολα συμβόλων εισόδου και εξόδου $\mathcal{X}$ και $\mathcal{Y}$, αντιστοίχως, και από ένα σύνολο (μία οικογένεια) δεσμευμένων συναρτήσεων μάζας πιθανότητας $p(y|x)$, μια για κάθε $x \in \mathcal{X}$, ώστε, για κάθε x και y, $p(y|x) \geq 0$ και, για κάθε x, $\sum_y p(y|x) = 1$. Η τ.μ. X είναι η είσοδος του καναλιού και η Y η έξοδός του.

Διακριτά κανάλια – Ορισμοί (2)

- Ορισμός 5.2. Ένα διακριτό κανάλι δεν έχει μνήμη (discrete memoryless channel - DMC) εάν

$$p(y_k|x^k, y^{k-1}, m) = p(y_k|x_k),$$

όπου $m \in \mathcal{M}$ είναι το μήνυμα που θέλουμε να μεταδώσουμε με n χρήσεις του καναλιού

- Παρατηρήστε ότι, από το Θεώρημα Ολικής Πιθανότητας, ο ορισμός αυτός συνεπάγεται και ότι

$$p(y_k|x^k, y^{k-1}) = p(y_k|x_k).$$

- Ορισμός 5.3. Έστω ότι χρησιμοποιούμε ένα διακριτό κανάλι n φορές. Ορίζουμε τη n-οστή επέκταση του διακριτού καναλιού χωρίς μνήμη ως $(\mathcal{X}^n, p(y^n|x^n), \mathcal{Y}^n)$, όπου

$$p(y_k|x^k, y^{k-1}) = p(y_k|x_k), \qquad k = 1, 2, \ldots, n.$$

Μετάδοση σε διακριτά κανάλια χωρίς ανάδραση

- Μπορούμε να γράψουμε

$$p(y^n|x^n, m) = \prod_{i=1}^{n} p(y_i|x^n, y^{i-1}, m).$$

- Εάν το κανάλι *χωρίς μνήμη* χρησιμοποιείται χωρίς ανάδραση, δηλαδή η είσοδος x^n στο κανάλι δεν εξαρτάται από τις εξόδους σε προηγούμενες χρονικές στιγμές αλλά μόνο από το μήνυμα m, $p(y_i|x^n, y^{i-1}, m) = p(y_i|x^i, y^{i-1}, m) = p(y_i|x_i)$ και

$$p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i).$$

($p(y_i|x^n, y^{i-1}, m) = p(y_i|y^{i-1}, m) = p(y_i|x^i, y^{i-1}, m)$ επειδή η αντιστοίχιση $m \leftrightarrow x^n$ είναι ένα-προς-ένα λόγω απουσίας ανάδρασης).