EE728
Προχωρημένα Θέματα Θεωρίας Πληροφορίας
8η διάλεξη

Δημήτρης-Αλέξανδρος Τουμπακάρης

Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών

21 Απριλίου 2010

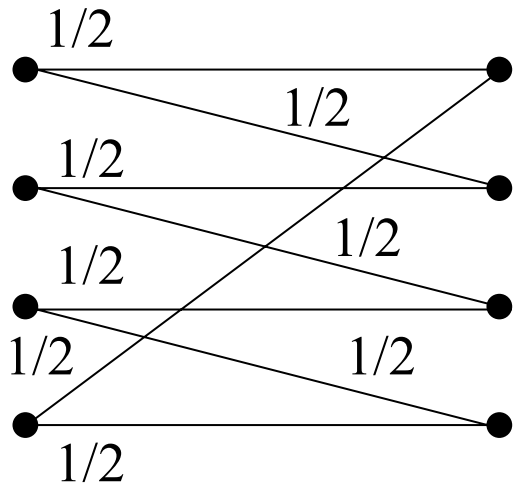
Περιεχόμενα σημερινού μαθήματος

Αντιστοιχία με συγγράμματα

- Cover & Thomas: 7.12, 7.13, 8.1–8.6
- Gallager: 5.1 – 5.6, 2.4, 7.1 – 7.2

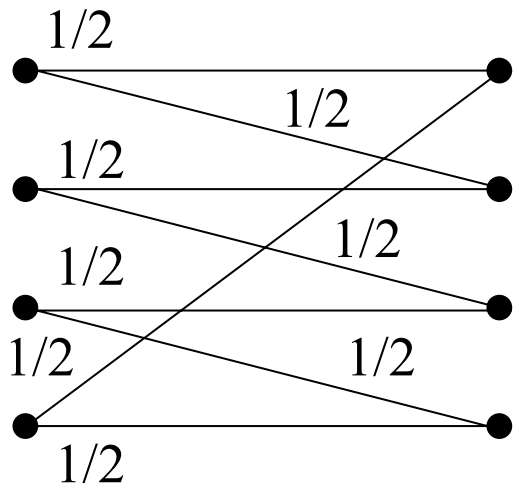
Χωρητικότητα καναλιών με ανάδραση

Παράδειγμα 8.1



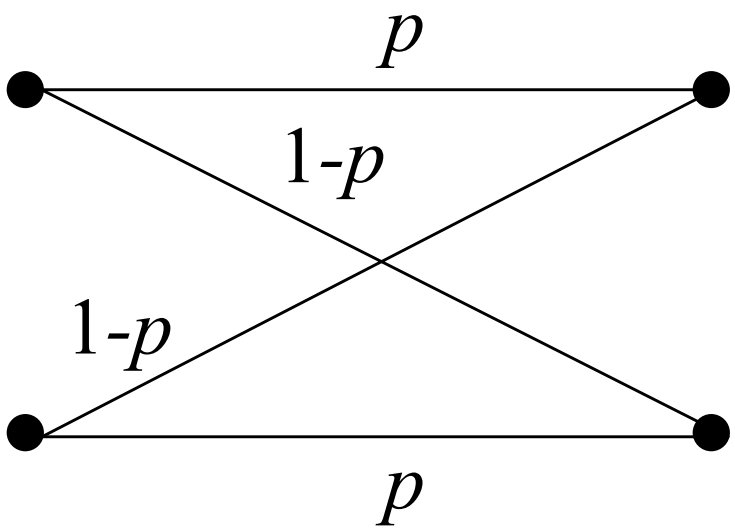
- Θεωρούμε το διακριτό κανάλι χωρίς μνήμη του σχήματος ("ενθόρυβη γραφομηχανή").
- Η χωρητικότητα του καναλιού ισούται με $C = \max I(X;Y) = \max\{H(Y) - H(Y|X)\} = 2 - 1 = 1$ bit.
- Μπορούμε να επιτύχουμε μετάδοση με ρυθμό ίσο με τη χωρητικότητα και με μηδενική πιθανότητα σφάλματος χρησιμοποιώντας π.χ. τις εισόδους 0 και 2. Προφανώς, $R = 1$ bit $= C$.

Παράδειγμα 8.1 (συνέχεια)



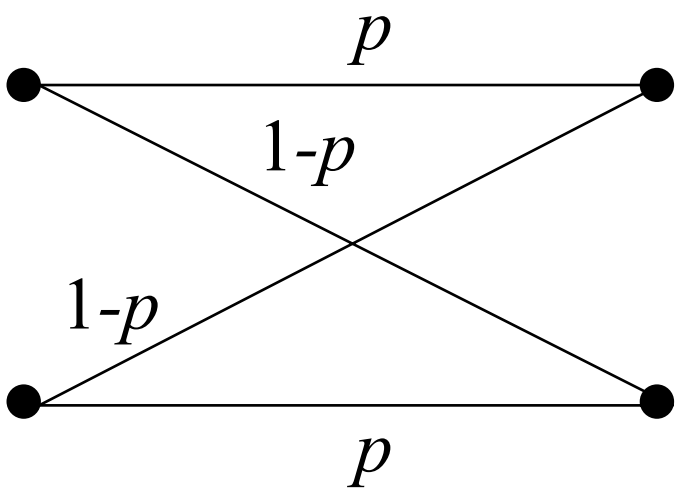
- Ό,τι και να συμβεί στο κανάλι είμαστε βέβαιοι ότι δε θα εμφανιστεί σφάλμα αποκωδικοποίησης.
- Εάν μπορούσαμε να χρησιμοποιήσουμε ανάδραση (feedback), η χωρητικότητα θα παρέμενε η ίδια;

Παράδειγμα 8.2



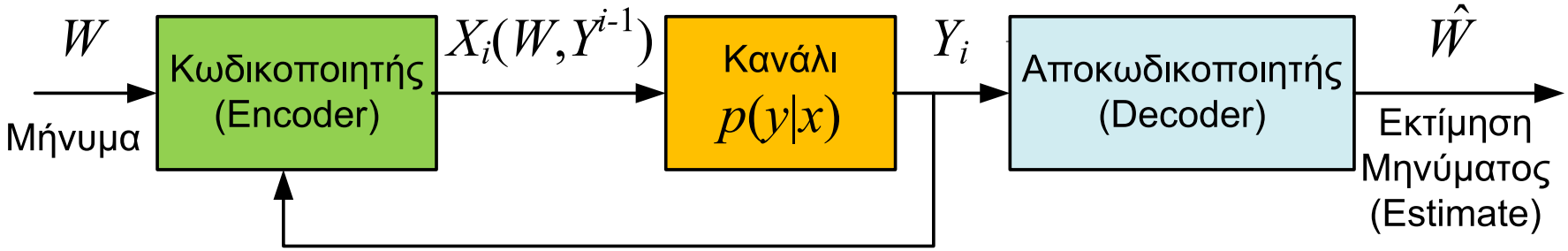
- Ας θεωρήσουμε, τώρα, το δυαδικό συμμετρικό κανάλι.
- Γνωρίζουμε ότι $C = 1 - H(p)$ και ότι η χωρητικότητα επιτυγχάνεται χρησιμοποιώντας και τα δύο μηνύματα με ίση πιθανότητα. Επομένως, κάθε φορά που στέλνουμε ένα από τα δύο μηνύματα στο κανάλι δε γνωρίζουμε εάν το μήνυμα μεταδόθηκε επιτυχώς. Η πιθανότητα σφάλματος ανά μετάδοση είναι μη μηδενική.

Παράδειγμα 8.2 (συνέχεια)



- Τι συμβαίνει εάν χρησιμοποιήσουμε ανάδραση; (όπου γνωρίζουμε εάν έχει εμφανιστεί σφάλμα στο δέκτη;)
- Σημείωση: Όταν χρησιμοποιούμε ανάδραση στο BSC, ο πομπός γνωρίζει ότι συνέβη σφάλμα, όχι, όμως, ο δέκτης!
- Παρόλο που κανείς θα περίμενε, ίσως, το αντίθετο, θα αποδείξουμε ότι, σε διακριτά κανάλια χωρίς μνήμη, η χρήση ανάδρασης δεν αυξάνει τη χωρητικότητα!

Χωρητικότητα καναλιού με ανάδραση – Μοντέλο



- Στο μοντέλο του σχήματος θεωρούμε ότι ο δέκτης στέλνει όλα τα ληφθέντα σύμβολα Y_i στον πομπό άμεσα και χωρίς σφάλματα. Ο πομπός χρησιμοποιεί την πληροφορία που λαμβάνει από το δέκτη προκειμένου να αποφασίσει πώς θα μεταδώσει.

Χωρητικότητα καναλιού με ανάδραση – Ορισμοί

Ορισμός Κώδικας ανάδρασης (feedback code) $(2^{nR}, n)$:

- Μια ακολουθία απεικονίσεων $x_i(W, Y^{i-1})$, όπου κάθε x_i είναι συνάρτηση του τρέχοντος μηνύματος W, καθώς και των σημάτων που ελήφθησαν στο δέκτη έως και τη χρονική στιγμή $i-1$: $Y_1, Y_2, \ldots, Y_{i-1}$ και
- Μια ακολουθία συναρτήσεων αποκωδικοποίησης $g : \mathcal{Y}^n \to \{1, 2, \ldots, 2^{nR}\}$.

- Θεωρούμε ότι τα μηνύματα W είναι ομοιόμορφα κατανεμημένα. Επομένως, $P_e^{(n)} = \Pr\{g(Y^n) \neq W\}$, όπου $X^n = X^n(W)$.

Ορισμός Η (λειτουργική) χωρητικότητα με ανάδραση, C_{FB}, του διακριτού καναλιού χωρίς μνήμη ισούται με το μέγιστο ρυθμό που είναι εφικτός με χρήση κωδίκων ανάδρασης.

Χωρητικότητα καναλιού με ανάδραση

Θεώρημα (Cover 7.12.1): $C_{FB} = C = \max_{p(x)} I(X;Y)$.

Απόδειξη Είναι, κατ' αρχάς, προφανές ότι $C_{FB} \geq C$, δεδομένου ότι το κανάλι χωρίς ανάδραση μπορεί να θεωρηθεί ως ειδική περίπτωση του καναλιού με ανάδραση.

- Θα αποδείξουμε ότι $C \geq C_{FB}$ και, επομένως, $C = C_{FB}$.
- Θα χρησιμοποιήσουμε και πάλι την ανισότητα Fano, όπως και στο αντίστροφο του Θεωρήματος Κωδικοποίησης Καναλιού. Ωστόσο, η απόδειξη διαφέρει γιατί στο κανάλι με ανάδραση δεν ισχύει η σχέση $I(X^n;Y^n) \leq nC$.

Χωρητικότητα καναλιού με ανάδραση (2)

- Υπενθυμίζεται ότι θεωρούμε ότι το μήνυμα W είναι ομοιόμορφα κατανεμημένο στο σύνολο $\{1, 2, \ldots, 2^{nR}\}$.

$$nR = H(W) \overset{(a)}{=} H(W|\hat{W}) + I(W;\hat{W}) \overset{(b)}{\leq} 1 + P_e^{(n)} nR + I(W;\hat{W})$$
$$\overset{(c)}{\leq} 1 + P_e^{(n)} nR + I(W;Y^n),$$

(a) Σχέση αμοιβαίας πληροφορίας-εντροπίας, (b) ανισότητα Fano, (c) ανισότητα επεξεργασίας δεδομένων.

Χωρητικότητα καναλιού με ανάδραση (3)

Η $I(W; Y^n)$ μπορεί να γραφτεί ως εξής:

$$I(W; Y^n) = H(Y^n) - H(Y^n|W) \overset{(a)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, Y_2, \ldots, Y_{i-1}, W)$$

$$\overset{(b)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, Y_2, \ldots, Y_{i-1}, W, X_i)$$

$$\overset{(c)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i)$$

$$\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) = \sum_{i=1}^{n} I(X_i; Y_i) \leq nC.$$

(a) Κανόνας αλυσίδας εντροπίας, (b) εάν γνωρίζουμε την ακολουθία Y^{i-1} και το μήνυμα W, γνωρίζουμε και το σύμβολο X_i που μεταδίδεται, (c) Το κανάλι δεν έχει μνήμη.

Χωρητικότητα καναλιού με ανάδραση (4)

- Επομένως, $I(W; Y^n) \leq nC$, και

$$nR \leq 1 + P_e^{(n)} nR + I(W; Y^n) \leq P_e^{(n)} nR + 1 + nC.$$

- Διαιρώντας με n, και για $n \to \infty$,

$$R \leq C, \text{ και, επομένως, } C_{FB} \leq C.$$

- Παρόλο που η χρήση ανάδρασης σε διακριτά κανάλια χωρίς μνήμη δεν αυξάνει τη χωρητικότητα, ενδέχεται να διευκολύνει τη μετάδοση. Για παράδειγμα, στο κανάλι διαγραφής, η μετάδοση απλουστεύεται εάν γνωρίζουμε πότε το σήμα εισόδου διαγράφεται.
- Φυσικά, στην πράξη, μπορεί να μην υπάρχει αξιόπιστος δίαυλος ανάδρασης ή να έχει κόστος (π.χ. σε εύρος ζώνης ή καθυστέρηση).

Αποκωδικοποίηση Μέγιστης Πιθανοφάνειας και Εκθέτης Σφάλματος

Αποκωδικοποίηση Μέγιστης εκ των Υστέρων Πιθανότητας (Maximum A Posteriori Probability -- MAP)

- Για την απόδειξη του Θεωρήματος Κωδικοποίησης Καναλιού υποθέσαμε ότι η αποκωδικοποίηση βασίζεται στην Ιδιότητα Από Κοινού Ασυμπτωτικής Ισοδιαμέρισης (Joint AEP).
- Δείξαμε ότι εάν η αποκωδικοποίηση βασίζεται στο Joint AEP μπορούμε να μεταδώσουμε με ρυθμούς αυθαίρετα κοντά στη χωρητικότητα με αυθαίρετα μικρή πιθανότητα σφάλματος.
- Αποδείξαμε ότι δεν μπορούμε να υπερβούμε τη χωρητικότητα. Επομένως, η αποκωδικοποίηση με χρήση από κοινού τυπικών ακολουθιών είναι ασυμπτωτικά βέλτιστη.

Αποκωδικοποίηση Μέγιστης εκ των Υστέρων Πιθανότητας (2)

- Εάν το κριτήριο είναι να ελαχιστοποιηθεί η πιθανότητα σφάλματος στο δέκτη, πρέπει να χρησιμοποιηθεί αποκωδικοποίηση Μέγιστης εκ των Υστέρων Πιθανότητας (Maximum a Posteriori (MAP) probability decoding).
- Θεωρούμε την πιθανότητα $p(y^n|x^n(w))$ να ληφθεί η ακολουθία y^n στο δέκτη δεδομένου ότι εστάλη ακολουθία $x^n(w)$ η οποία αντιστοιχεί στο μήνυμα w (η κωδική λέξη του μηνύματος w).

Αποκωδικοποίηση Μέγιστης εκ των Υστέρων Πιθανότητας (3)

- Από τον κανόνα του Bayes,

$$p(w|y^n) = \frac{p(y^n|x^n(w))p(w)}{p(y^n)}, \text{ όπου } p(y^n) = \sum_{w=1}^{|\mathcal{W}|} p(w)p(y^n|x^n(w)).$$

- Εάν ο δέκτης αποκωδικοποιεί την ακολουθία y^n στο μήνυμα w, η πιθανότητα σφάλματος δεδομένης της ληφθείσας ακολουθίας y^n ισούται με $1 - p(w|y^n)$.
- Επομένως, για να ελαχιστοποιηθεί η πιθανότητα σφάλματος, πρέπει να επιλεγεί το μήνυμα w το οποίο μεγιστοποιεί την εκ των υστέρων (a posteriori) πιθανότητα του w δεδομένης της ληφθείσας ακολουθίας y^n ($p(w|y^n)$).

Κανόνας αποκωδικοποίησης MAP

Κανόνας αποκωδικοποίησης MAP

$\hat{w} = g(y^n)$, τέτοιο ώστε

$$p(\hat{w}|y^n) \geq p(w'|y^n), \text{ για όλα τα } w' \neq \hat{w}, \ \hat{w}, w' \in \mathcal{W}$$

Εναλλακτική έκφραση

$$\hat{w} = g(y^n) = \arg\max_{w'} p(w'|y^n)$$

Αποκωδικοποίηση Μέγιστης εκ των Υστέρων Πιθανότητας (4)

- Με χρήση του κανόνα του Bayes,

$$p(w|y^n) \geq p(w'|y^n) \Rightarrow$$
$$\frac{p(y^n|x^n(w))p(w)}{p(y^n)} \geq \frac{p(y^n|x^n(w'))p(w')}{p(y^n)}$$

- Επομένως, ο κανόνας MAP μπορεί να γραφτεί ως:

Κανόνας MAP

$$p(y^n|x^n(w))p(w) \geq p(y^n|x^n(w'))p(w')$$

- Για κανάλι χωρίς μνήμη,

Κανόνας MAP για κανάλι χωρίς μνήμη

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

Αποκωδικοποίηση Μέγιστης Πιθανοφάνειας (Maximum Likelihood (ML) decoding)

- Με βάση τον κανόνα MAP επιλέγεται το μήνυμα που ικανοποιεί τη σχέση $p(y^n|x^n(w))p(w) \geq p(y^n|x^n(w'))p(w')$ για όλα τα $w' \neq w$.
- Εάν όλα τα μηνύματα εκπέμπονται με την ίδια πιθανότητα (ομοιόμορφα), ο αποκωδικοποιητής μπορεί να αποκωδικοποιήσει με βάση τη σχέση

$$p(y^n|x^n(w)) \geq p(y^n|x^n(w')) \text{ για όλα τα } w' \neq w.$$

- Η αποκωδικοποίηση με βάση την παραπάνω σχέση ονομάζεται μέγιστης πιθανοφάνειας. Στη γενική περίπτωση (όπου τα μηνύματα δεν ακολουθούν ομοιόμορφη κατανομή) δε μεγιστοποιεί την πιθανότητα να έχει μεταδοθεί το μήνυμα w δεδομένης της ακολουθίας y^n.
- Ωστόσο, μεγιστοποιείται η πιθανότητα να έχει ληφθεί η y^n δεδομένου του w.

Γιατί ML και όχι MAP;

- Στη γενική περίπτωση (όπου η κατανομή των μηνυμάτων στην είσοδο του καναλιού δεν είναι ομοιόμορφη) η αποκωδικοποίηση ML δεν είναι βέλτιστη.
- Ωστόσο, στην πράξη, η αποκωδικοποίηση ML χρησιμοποιείται συχνότερα από την αποκωδικοποίηση MAP. Κάποιοι από τους λόγους είναι οι εξής:
  - Πολύ συχνά, τα μηνύματα που στέλνονται είναι ισοπίθανα (π.χ. όταν έχει γίνει καλή συμπίεση πριν από τη μετάδοση), οπότε η αποκωδικοποίηση ML είναι βέλτιστη.
  - Αποδεικνύεται (βλ. π.χ. Cioffi, http://www.stanford.edu/group/cioffi/book/chap1.pdf) ότι, εάν η κατανομή των μηνυμάτων $p(w)$ είναι άγνωστη, η αποκωδικοποίηση ML ελαχιστοποιεί την πιθανότητα σφάλματος για τη "χειρότερη" κατανομή εισόδου.
- Πολλές φορές η αποκωδικοποίηση ML είναι πολύπλοκη, οπότε χρησιμοποιούνται υποβέλτιστες μέθοδοι. Περισσότερα στα μαθήματα Ψηφιακών Επικοινωνιών.

Εκθέτης Σφάλματος (Error Exponent) (εισαγωγή)

- Σύμφωνα με το Θεώρημα Κωδικοποίησης Καναλιού, είναι δυνατόν να μεταδώσουμε σε διακριτό κανάλι χωρίς μνήμη με αυθαίρετα μικρή πιθανότητα σφάλματος, εφόσον ο ρυθμός μετάδοσης δεν υπερβαίνει τη χωρητικότητα.
- Αντίστροφα, δεν υπάρχει κώδικας με αυθαίρετα μικρή πιθανότητα σφάλματος ο οποίος επιτυγχάνει μετάδοση με ρυθμό μεγαλύτερο από τη χωρητικότητα καναλιού.
- Αποδείξαμε το Θεώρημα Κωδικοποίησης Καναλιού όταν ο δέκτης αποκωδικοποιεί με βάση την Ιδιότητα Από Κοινού Ασυμπτωτικής Ισοδιαμέρισης. Το Θεώρημα αποδεικνύεται και για αποκωδικοποίηση μέγιστης πιθανοφάνειας (ML – βλ. π.χ. Gallager).

Εκθέτης Σφάλματος (Error Exponent) (2)

- Στην απόδειξη, για να επιτύχουμε αυθαίρετα μικρή πιθανότητα σφάλματος, αφήσαμε το n να τείνει στο άπειρο.
- Τι συμβαίνει όταν το n είναι πεπερασμένο; Πώς μεταβάλλεται η πιθανότητα σφάλματος ως συνάρτηση του n;
- Ένας τρόπος να ποσοτικοποιηθεί η εξάρτηση της μέσης πιθανότητας σφάλματος από το n είναι ο εκθέτης σφάλματος (error exponent) ο οποίος παρέχει ένα άνω φράγμα όταν χρησιμοποιείται αποκωδικοποίηση μέγιστης πιθανοφάνειας.

Εκθέτης Σφάλματος (Error Exponent) (3)

Θεώρημα (Gallager 5.6.2 & Corollary 1): Έστω διακριτό κανάλι χωρίς μνήμη με πίνακα μετάβασης $p(y_j|x_k), j = 1, \ldots, J$ και $k = 1, \ldots, K$. Για δεδομένο n και R θεωρούμε το σύνολο των κωδίκων $(2^{nR}, n)$ των οποίων τα σύμβολα επιλέγονται ανεξάρτητα με βάση κατανομή $p(x)$. Εάν ο δέκτης χρησιμοποιεί αποκωδικοποίηση μέγιστης πιθανοφάνειας, για τη μέση τιμή σφάλματος υπολογισμένη για όλους τους τυχαίους κώδικες οι οποίοι παράγονται με βάση κατανομή $p^*(x)$ και για όλα τα πιθανά μηνύματα, ισχύει

$$P_e^{(n)} \leq \exp\{-nE_r(R)\},$$

όπου $E_r(R)$ είναι ο εκθέτης τυχαίας κωδικοποίησης ή εκθέτης σφάλματος (random coding/error exponent)

$$E_r(R) = \max_{0 \leq \rho \leq 1} \max_{p(x)} \{E_0(\rho, p(x)) - \rho R\},$$

$p^*(x)$ η κατανομή που επιτυγχάνει τον $E_r(R)$ και

$$E_0(\rho, p(x)) = -\log \sum_{j=1}^{J} \left[\sum_{k=1}^{K} p(x_k) p(y_j|x_k)^{1/(1+\rho)} \right]^{1+\rho}.$$

Εκθέτης Σφάλματος (Error Exponent) (4)

- Παρόλο που η έκφραση για τον εκθέτη σφάλματος είναι σχετικά πολύπλοκη, βασίζεται σε απλά βήματα (βλ. Gallager 5.6).
- Εάν μπορούμε να υπολογίσουμε τον $E_r(R)$ για δεδομένο διακριτό κανάλι χωρίς μνήμη, αποκτούμε ένα φράγμα για την πιθανότητα σφάλματος για δεδομένο ρυθμό μετάδοσης και δεδομένο μήκος κώδικα n: $P_e^{(n)} \leq \exp\{-nE_r(R)\}$.
- Αποδεικνύεται ότι, για $0 \leq R < C, E_r(R) > 0$ και, επομένως, με κατάλληλη κωδικοποίηση, η πιθανότητα σφάλματος μπορεί να κρατηθεί αυθαίρετα κοντά στο μηδέν με χρήση κωδίκων κατάλληλου μήκους n.
- Όπως και στην περίπτωση αποκωδικοποίησης με χρήση από κοινού τυπικότητας, το γεγονός ότι $P_e^{(n)} \leq \exp\{-nE_r(R)\}$ δε συνεπάγεται ότι η πιθανότητα σφάλματος $P_{e,w}^{(n)}$ που αντιστοιχεί στην κωδική λέξη $x^n(w)$ θα είναι $\leq \exp\{-nE_r(R)\}$. Ωστόσο, αποδεικνύεται (Gallager 5.6 Corollary 2) ότι υπάρχει κώδικας $(2^{nR}, n)$ τέτοιος ώστε $P_{e,w}^{(n)} \leq 4\exp\{-nE_r(R)\}$ για όλα τα $w = 1, 2, \ldots, 2^{nR}$.

Θεώρημα Διαχωρισμού Πηγής - Καναλιού

Θεώρημα Διαχωρισμού Πηγής - Καναλιού – Εισαγωγή

- Γνωρίζουμε, πλέον, ότι για να συμπιέσουμε μια πηγή με ρυθμό εντροπίας $H(\mathcal{X})$ χρειαζόμαστε $R > H(\mathcal{X})$ bits/σύμβολο.
- Αντίστοιχα, για να μεταδώσουμε R bits/χρήση καναλιού μέσω διακριτού καναλιού χωρίς μνήμη πρέπει $R < C$.
- Έστω ότι θέλουμε να μεταδώσουμε τα μηνύματα πηγής με ρυθμό εντροπίας $H(\mathcal{X})$ με χρήση καναλιού χωρητικότητας C. Είναι η συνθήκη $H(\mathcal{X}) < C$ ικανή και αναγκαία για να μπορεί να γίνει μετάδοση των μηνυμάτων της πηγής;

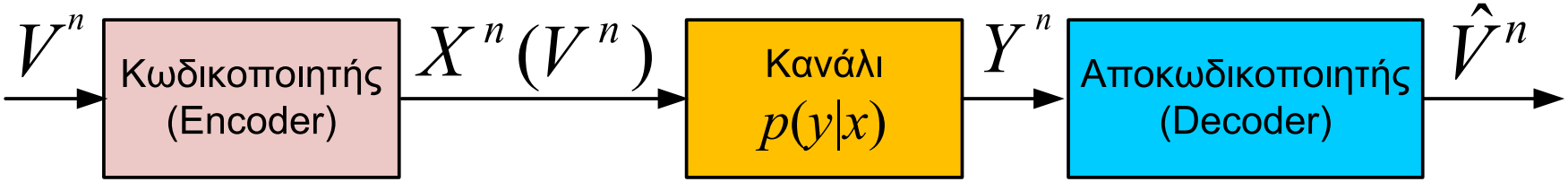
Θεώρημα Διαχωρισμού Πηγής - Καναλιού – Εισαγωγή (2)

- Ειδικότερα, είναι βέλτιστο να συμπιέσουμε την πηγή κοντά στο ρυθμό εντροπίας της και μετά να μεταδώσουμε τη συμπιεσμένη ακολουθία στο κανάλι ή μήπως υπάρχει πιο αποδοτικός τρόπος μετάδοσης (και, άρα, τρόπος να μεταδώσουμε με μεγαλύτερο ρυθμο ;)
- Θα αποδείξουμε ότι η μετάδοση με συμπίεση της πηγής και, στη συνέχεια, με κωδικοποίηση καναλιού είναι το ίδιο αποδοτική με οποιαδήποτε άλλη μέθοδο. Δηλαδή, εφόσον $H(\mathcal{X}) < C$, μπορούμε να συμπιέσουμε την πηγή και να μεταδώσουμε την πληροφορία που παράγει μέσω του καναλιού. Αντίστροφα, προκειμένου η πληροφορία μιας πηγής να μεταδίδεται με αυθαίρετα μικρή πιθανότητα σφάλματος στο κανάλι, πρέπει να ισχύει $H(\mathcal{X}) < C$.

Θεώρημα Διαχωρισμού Πηγής - Καναλιού – Εισαγωγή (3)

- Παρόλο που το Θεώρημα Διαχωρισμού Πηγής - Καναλιού φαίνεται προφανές, υπάρχουν περιπτώσεις στις οποίες δεν ισχύει! (κανάλια πολλών χρηστών).
- Στις περιπτώσεις που το Θεώρημα ισχύει, διευκολύνεται ο σχεδιασμός Συστημάτων Επικοινωνιών, δεδομένου ότι ο Κωδικοποιητής Πηγής και ο Κωδικοποιητής Καναλιού μπορούν να σχεδιαστούν ανεξάρτητα. Για παράδειγμα, ο τρόπος μετάδοσης σε μια γραμμή ADSL ή σε ένα δίκτυο WiFi είναι ο ίδιος, ανεξάρτητα από το εάν ο χρήστης στέλνει μουσική ή εικόνες ή κείμενο.
- Ωστόσο, το γεγονός ότι η μέθοδος δύο βημάτων που συνίσταται στη συμπίεση της πηγής ανεξάρτητα από το κανάλι και στη μετάδοση της συμπιεσμένης ακολουθίας δε συνεπάγεται απώλειες, δε σημαίνει, κατ' ανάγκη, ότι είναι πάντοτε και η λιγότερο πολύπλοκη.

Θεώρημα Διαχωρισμού Πηγής - Καναλιού



- Θεωρούμε πηγή V η οποία παράγει σύμβολα από πεπερασμένο αλφάβητο $\mathcal{V}$. Η πηγή ικανοποιεί τη (γενικευμένη) Ιδιότητα Ασυμπτωτικής Ισοδιαμέρισης αλλά δεν είναι, κατ' ανάγκη, χωρίς μνήμη. Στη γενική περίπτωση είναι στάσιμη και εργοδική.
- Ο πομπός απεικονίζει την ακολουθία $V^n = V_1, V_2, \ldots, V_n$ της πηγής σε κωδική λέξη $X^n(V^n)$ και τη μεταδίδει στο κανάλι.
- Ο δέκτης παράγει εκτίμηση $\hat{V}^n$ της μεταδοθείσας ακολουθίας με βάση τη ληφθείσα ακολουθία Y^n. Όταν $\hat{V}^n \neq V^n$ εμφανίζεται σφάλμα στο δέκτη.

Θεώρημα Διαχωρισμού Πηγής - Καναλιού (συνέχεια)

- Η πιθανότητα σφάλματος ισούται με

$$\Pr\{V^n \neq \hat{V}^n\} = \sum_{y^n} \sum_{v^n} p(v^n) p(y^n | x^n(v^n)) I(g(y^n) \neq v^n),$$

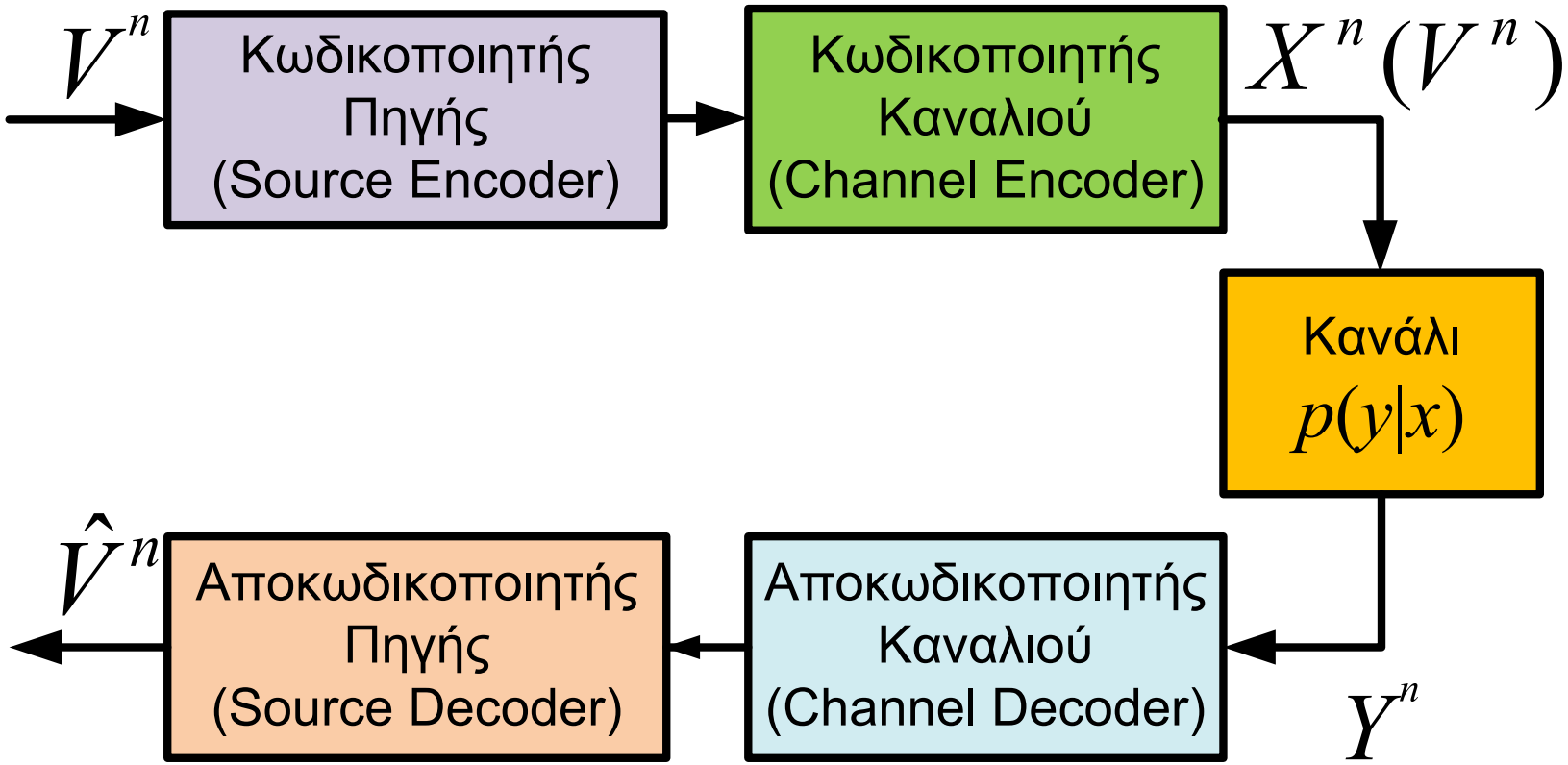
όπου I η συνάρτηση-δείκτης και $g(\cdot)$ η συνάρτηση αποκωδικοποίησης.

- Θεώρημα Διαχωρισμού Πηγής - Καναλιού (ευθύ): Έστω $V_1, V_2, \ldots, V_n$ στοχαστική ανέλιξη με πεπερασμένο αλφάβητο η οποία ικανοποιεί το AEP, και για την οποία ισχύει $H(\mathcal{V}) < C$. Υπάρχει κώδικας πηγής-καναλιού με πιθανότητα σφάλματος $\Pr\{\hat{V}^n \neq V^n\} \to 0$.

- Αντίστροφα, για κάθε στάσιμη και εργοδική στοχαστική ανέλιξη, εάν $H(\mathcal{V}) > C$, η πιθανότητα σφάλματος δεν μπορεί να βρίσκεται αυθαίρετα κοντά στο 0 και, εμπομένως, δεν είναι δυνατή η μετάδοση της στοχαστικής ανέλιξης μέσω του καναλιού με αυθαίρετα μικρή πιθανότητα σφάλματος.

Θεώρημα Διαχωρισμού Πηγής - Καναλιού
Απόδειξη ευθέος

Θα χρησιμοποιήσουμε κωδικοποίηση δύο φάσεων:
1) Κωδικοποίηση πηγής (συμπίεση) και 2) Κωδικοποίηση καναλιού.



Θεώρημα Διαχωρισμού Πηγής - Καναλιού
Απόδειξη ευθέος (2)

- Από το AEP, για μεγάλο n το τυπικό σύνολο περιέχει $\leq 2^{n(H(\mathcal{V})+\epsilon)}$ στοιχεία και σχεδόν όλη την πιθανότητα. Κωδικοποιούμε μόνο τις τυπικές ακολουθίες και αγνοούμε τις υπόλοιπες. Επομένως, χρειαζόμαστε το πολύ $n(H(\mathcal{V})+\epsilon)$ bits.
- Προκειμένου να μεταδώσουμε τα $n(H(\mathcal{V})+\epsilon)$ bits στο κανάλι πρέπει να ισχύει
$$H(\mathcal{V})+\epsilon = R < C.$$
- Ο δέκτης αποκωδικοποιεί με βάση την από κοινού τυπικότητα. Για την πιθανότητα σφάλματος ισχύει
$$\Pr\{V^n \neq \hat{V}^n\} \leq \Pr\{V^n \notin A_\epsilon^{(n)}\} + \Pr\{g(Y^n) \neq V^n | V^n \in A_\epsilon^{(n)}\}.$$

Θεώρημα Διαχωρισμού Πηγής - Καναλιού
Απόδειξη ευθέος (3)

$$\Pr\{V^n \neq \hat{V}^n\} \leq \Pr\{V^n \notin A_\epsilon^{(n)}\} + \Pr\{g(Y^n) \neq V^n | V^n \in A_\epsilon^{(n)}\}.$$

- Για αρκούντως μεγάλο n, από το AEP, $\Pr\{V^n \notin A_\epsilon^{(n)}\} \leq \epsilon$.
- Ομοίως, από το Joint AEP, για αρκούντως μεγάλο n, και δεδομένου ότι $H(\mathcal{V}) + \epsilon = R < C$, $\Pr\{g(Y^n) \neq V^n | V^n \in A_\epsilon^{(n)}\} \leq \epsilon$.
- Συνεπώς, για οποιοδήποτε ϵ, και εφόσον $H(\mathcal{V}) + \epsilon < C$, υπάρχει μήκος κωδικής λέξης n_0 τέτοιο ώστε, για $n > n_0$, $\Pr\{V^n \neq \hat{V}^n\} \leq 2\epsilon$.
- Επομένως, χρησιμοποιώντας τη μέθοδο δύο φάσεων, μπορούμε να μεταδώσουμε με αυθαίρετα μικρή πιθανότητα σφάλματος εφόσον $H(\mathcal{V}) < C$.

Θεώρημα Διαχωρισμού Πηγής - Καναλιού
Απόδειξη αντιστρόφου

- Θα δείξουμε ότι, για οποιαδήποτε μέθοδο κωδικοποίησης (ακόμα και τυχαία) $X^n(V^n) : \mathcal{V}^n \to \mathcal{X}^n$ και αποκωδικοποίησης $g(Y^n) : \mathcal{Y}^n \to \mathcal{V}^n$, εάν $\Pr\{\hat{V}^n \neq V^n\} \to 0$, τότε $H(\mathcal{V}) \leq C$.
- Από την ανισότητα Fano,

$$H(V^n|\hat{V}^n) \leq 1 + \Pr\{\hat{V}^n \neq V^n\} \log |\mathcal{V}^n| = 1 + n \Pr\{\hat{V}^n \neq V^n\} \log |\mathcal{V}|.$$

- Θα υπολογίσουμε άνω φράγμα για την $H(\mathcal{V})$

$$\begin{aligned} H(\mathcal{V}) &\overset{(a)}{\leq} \frac{H(V_1, V_2, \ldots, V_n)}{n} = \frac{H(V^n)}{n} = \frac{1}{n} H(V^n|\hat{V}^n) + \frac{1}{n} I(V^n; \hat{V}^n) \\ &\overset{(b)}{\leq} \frac{1}{n} \left(1 + n \Pr\{\hat{V}^n \neq V^n\} \log |\mathcal{V}|\right) + \frac{1}{n} I(V^n; \hat{V}^n) \end{aligned}$$

(a) Ρυθμός εντροπίας για στάσιμες στοχαστικές ανελίξεις, (b) Ανισότητα Fano

Θεώρημα Διαχωρισμού Πηγής - Καναλιού
Απόδειξη αντιστρόφου (συνέχεια)

$$
\begin{aligned}
H(\mathcal{V}) &\leq \frac{1}{n}\left(1 + n\Pr\{\hat{V}^n \neq V^n\}\log|\mathcal{V}|\right) + \frac{1}{n}I(V^n;\hat{V}^n) \\
&\overset{(a)}{\leq} \frac{1}{n}\left(1 + n\Pr\{\hat{V}^n \neq V^n\}\log|\mathcal{V}|\right) + \frac{1}{n}I(X^n;Y^n) \\
&\overset{(b)}{\leq} \frac{1}{n} + \Pr\{\hat{V}^n \neq V^n\}\log|\mathcal{V}| + C.
\end{aligned}
$$

(a) Ανισότητα Επεξεργασίας Δεδομένων, (b) το κανάλι δεν έχει μνήμη.

- Για $n \to \infty$, $\Pr\{\hat{V}^n \neq V^n\} \to 0$ και, επομένως,

$$H(\mathcal{V}) \leq C.$$

Συνεχείς τ.μ. και Διαφορική Εντροπία

Διαφορική Εντροπία – Εισαγωγή

- Έως τώρα θεωρούσαμε διακριτές τ.μ. με τιμές με πεπερασμένο και διακριτό αλφάβητο.
- Τα αποτελέσματα της Θεωρίας Πληροφορίας εφαρμόζονται και για συνεχείς τ.μ., με κατάλληλες τροποποιήσεις και με χρήση της διαφορικής εντροπίας (differential entropy).
- Γενικά, όσα ισχύουν για διακριτές τ.μ. ισχύουν (με κατάλληλες τροποποιήσεις) και για συνεχείς τ.μ. Επομένως, θα αναφερθούμε στις συνεχείς τ.μ. πιο επιγραμματικά, φροντίζοντας, όμως, να επισημαίνουμε τις διαφορές, όπου υπάρχουν.

Διαφορική Εντροπία – Ορισμός

Ορισμός Η Διαφορική Εντροπία $h(X)$ συνεχούς τ.μ. X με συνάρτηση πυκνότητας πιθανότητας $f(x)$, εάν η f υπάρχει, ορίζεται ως

$$h(X) = -\int_S f(x)\log f(x)dx,$$

όπου S είναι το πεδίο ορισμού της τ.μ.

- Υποθέτουμε ότι η $f(x)\log f(x)$ είναι ολοκληρώσιμη κατά Riemann.

Παράδειγμα 8.3. – Δεν ισχύει, κατ' ανάγκη, $h(X) \geq 0$!

- Έστω συνεχής τ.μ. X, ομοιόμορφα κατανεμημένη στο διάστημα $[0, a]$.

$$h(X) = -\int_S f(x) \log f(x) dx = -\int_0^a \frac{1}{a} \log \frac{1}{a} dx = \log a.$$

- Για $a < 1$, $h(X) < 0$.
- Ωστόσο, η ποσότητα $2^{h(X)}$ είναι πάντοτε μη αρνητική.
- Η διαφορική εντροπία διακριτής τ.μ. ισούται με $-\infty$ ($2^{-\infty} = 0$).

Παράδειγμα 8.4. – Εντροπία Γκαουσιανής τ.μ.

- Έστω συνεχής τ.μ. X η οποία ακολουθεί Γκαουσιανή κατανομή με μέση τιμή 0 και διασπορά σ^2.

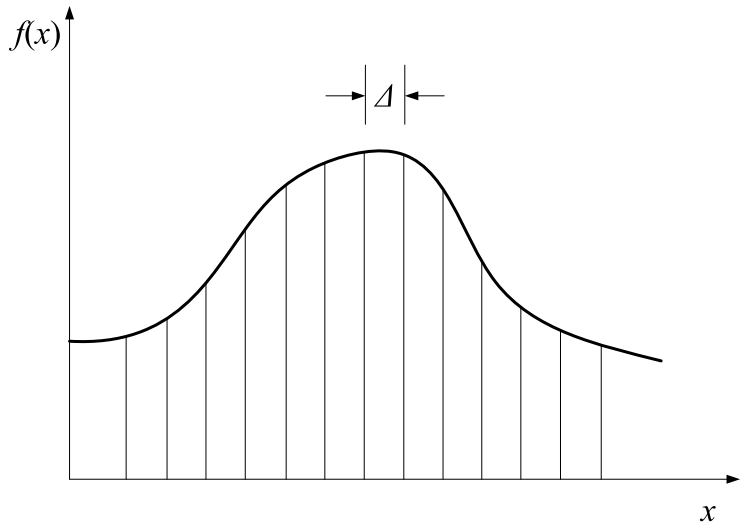
$$X \sim \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}.$$

- Με χρήση του ορισμού της διαφορικής εντροπίας,

$$\begin{aligned} h(X) &= -\int_S f(x)\ln f(x)dx = -\int_{-\infty}^{\infty} f(x)\ln f(x)dx \\ &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \left[-\frac{x^2}{2\sigma^2} - \ln\left(\sqrt{2\pi}\sigma\right)\right] dx \\ &= \frac{EX^2}{2\sigma^2} + \frac{1}{2}\ln 2\pi\sigma^2 = \frac{1}{2} + \frac{1}{2}\ln 2\pi\sigma^2 = \frac{1}{2}\ln e + \frac{1}{2}\ln 2\pi\sigma^2 \\ &= \frac{1}{2}\ln 2\pi e\sigma^2 \text{ nats} = \frac{1}{2}\log 2\pi e\sigma^2 \text{ bits} \end{aligned}$$

Παράδειγμα 8.5. – Εντροπία διακριτής αναπαράστασης συνεχούς τ.μ.

- Έστω συνεχής τ.μ. X με συνάρτηση πυκνότητας πιθανότητας $f(x)$. Χωρίζουμε την $f(X)$ σε κομμάτια πλάτους Δ, όπως φαίνεται στο Σχήμα.



- Για κάθε διάστημα πλάτους Δ υπάρχει x_i τέτοιο ώστε $f(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} f(x)dx$.
- Θεωρούμε τη διακριτή αναπαράσταση, X^Δ, της συνεχούς τ.μ. X:

$$X^\Delta = x_i, \quad \text{όταν } i\Delta \leq X < (i+1)\Delta.$$

Παράδειγμα 8.5. – Εντροπία διακριτής αναπαράστασης συνεχούς τ.μ. (2)

- $p_i \triangleq \Pr\{X^\Delta = x_i\} = \int_{i\Delta}^{(i+1)\Delta} f(x)dx = f(x_i)\Delta.$
- Επομένως, για την εντροπία της (διακριτής) X^Δ ισχύει

$$H(X^\Delta) = -\sum_{-\infty}^{\infty} p_i \log p_i = -\sum_{-\infty}^{\infty} (f(x_i)\Delta)\log\left(f(x_i)\Delta\right) =$$

$$= -\sum_{-\infty}^{\infty} f(x_i)\Delta \log f(x_i) - \sum_{-\infty}^{\infty} f(x_i)\Delta \log \Delta$$

$$= -\sum_{-\infty}^{\infty} f(x_i)\Delta \log f(x_i) - \log \Delta.$$

Παράδειγμα 8.5. – Εντροπία διακριτής αναπαράστασης συνεχούς τ.μ. (3)

$$H(X^{\Delta}) = -\sum_{-\infty}^{\infty} f(x_i)\Delta \log f(x_i) - \log \Delta.$$

- Όταν $\Delta \to 0$, $H(X^{\Delta}) \to h(X) - \log \Delta$, εφόσον η $f(x)$ είναι ολοκληρώσιμη κατά Riemann.
- Παρατηρούμε ότι η ποσότητα $\log \Delta$ είναι ανάλογη του αριθμού n των bits που χρησιμοποιούνται για τη διακριτοποίηση (κβαντισμό) της συνεχούς τ.μ. X. Επομένως, $H(X^{\Delta}) \approx h(X) + n$.
- Η ακριβής (μη κβαντισμένη) τιμή συνεχούς τ.μ. απαιτεί άπειρα bits για την περιγραφή της (διαισθητικά λογικό).

Από κοινού και υπό συνθήκη Διαφορική Εντροπία

Οι ορισμοί είναι παρόμοιοι με αυτούς για διακριτές τ.μ.

Ορισμός Από κοινού διαφορική εντροπία:
$h(X_1, X_2, \ldots, X_n) = -\int f(x^n) \log f(x^n) dx^n,$
όπου $f(x^n) = f(x_1, x_2, \ldots, x_n)$.

Ορισμός Υπό συνθήκη διαφορική εντροπία:
$h(X|Y) = -\int f(x, y) \log f(x|y) dxdy.$

- Όπως και στην περίπτωση διακριτών τ.μ., εάν όλες οι ποσότητες είναι πεπερασμένες,

$$h(X|Y) = h(X, Y) - h(Y).$$

Παράδειγμα 8.6 – Διαφορική Εντροπία πολυμεταβλητής Γκαουσιανής τ.μ.

- Έστω τ.μ. που ακολουθεί πολυμεταβλητή Γκαουσιανή κατανομή:

$$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T K^{-1} (\mathbf{x}-\mathbf{m})},$$

  όπου $(\cdot)^T$ υποδηλώνει αναστροφή (διανύσματος ή πίνακα), K είναι ο πίνακας συσχέτισης και $|K|$ η ορίζουσα του K.

- Αποδεικνύεται (με χρήση του ορισμού και πράξεις – Cover Theorem 8.4.1) ότι

$$h(X_1, X_2, \ldots, X_n) = h(\mathcal{N}_n(\mathbf{m}, K)) = \frac{1}{2} \log(2\pi e)^n |K| \text{bits}.$$

- Για πραγματική τ.μ. $X \sim \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mathbf{m})^2}{2\sigma^2}}$, $h(X) = \frac{1}{2} \log(2\pi e \sigma^2)$ bits.

Σχετική Εντροπία και Αμοιβαία Πληροφορία για συνεχείς τ.μ.

- Σχετική Εντροπία (Απόσταση Kullback-Leibler): $D(f||g) = \int f \log \frac{f}{g}$. Πεπερασμένη μόνο εφόσον το πεδίο ορισμού της f περιέχεται στο πεδίο ορισμού της g.
- Εάν ορίζεται από κοινού συνάρτηση πυκνότητας πιθανότητας για τις τ.μ. X και Y, η Αμοιβαία Πληροφορία ορίζεται ως

$$I(X;Y) = D(f(x,y)||f(x)f(y)) = \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} dxdy.$$

- Όπως και για τις διακριτές τ.μ., $I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X) = h(X) + h(Y) - h(X,Y)$.

Σχετική Εντροπία και Αμοιβαία Πληροφορία για συνεχείς τ.μ. (2)

- Εάν δεν ορίζεται $f(x, y)$, μπορεί να χρησιμοποιηθεί ο πλέον γενικός ορισμός της Αμοιβαίας Πληροφορίας

$$I(X;Y) = \sup_{\text{όλες οι } \mathcal{P},\mathcal{Q}} I([X]_{\mathcal{P}}; [Y]_{\mathcal{Q}}),$$

όπου $\mathcal{P}$ και $\mathcal{Q}$ πεπερασμένες διαμερίσεις (partitions) των $\mathcal{X}$ και $\mathcal{Y}$ και $[X]_{\mathcal{P}}$, $[Y]_{\mathcal{Q}}$ οι κβαντίσεις των X και Y ως προς τις διαμερίσεις $\mathcal{P}$ και $\mathcal{Q}$, αντίστοιχα (περισσότερες λεπτομέρειες στο βιβλίο των Cover & Thomas).

Ιδιότητες Σχετικής Εντροπίας και Αμοιβαίας Πληροφορίας για συνεχείς τ.μ.

- $D(f||g) \geq 0$, με ισότητα όταν $f = g$ σχεδόν παντού.
  Απόδειξη: Εάν S είναι το πεδίο ορισμού της f,

$$-D(f||g) = \int_S f \log \frac{g}{f} \overset{(a)}{\leq} \log \int_S f \frac{g}{f} = \log \int_S g \overset{(b)}{\leq} \log 1 = 0.$$

  (a) γιατί; (b) S υποσύνολο του πεδίου ορισμού της g.
- $I(X;Y) \geq 0$ με $=$ εάν και μόνο εάν X και Y ανεξάρτητες. Γιατί;

Ιδιότητες Σχετικής Εντροπίας και Αμοιβαίας Πληροφορίας για συνεχείς τ.μ. (2)

- $h(X|Y) \leq h(X)$ με $=$ εάν και μόνο εάν X και Y ανεξάρτητες.
- Κανόνας αλυσίδας για τη Διαφορική Εντροπία: $h(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} h(X_i|X_1, X_2, \ldots, X_{i-1})$. Αποδεικνύεται εύκολα από τον ορισμό της Από Κοινού Διαφορικής Εντροπίας.
- $h(X_1, X_2, \ldots, X_n) \leq \sum h(X_i)$, με $=$ εάν και μόνο εάν οι $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες.

Άλλες Ιδιότητες Διαφορικής Εντροπίας

- $h(X+c) = h(X)$. Προκύπτει απευθείας από τον ορισμό.
  - Η διαφορική εντροπία είναι αναλλοίωτη σε μετάθεση.
  - Αντίστοιχη ιδιότητα για διακριτές τ.μ.: η εντροπία διακριτών τ.μ. εξαρτάται μόνο από την κατανομή τους και όχι από τις τιμές τους.
- $h(aX) = h(X) + \log|a|$. Για την απόδειξη δείτε π.χ. Cover & Thomas Theorem 8.6.4.
  - Διαισθητικά λογικό: Η τ.μ. παίρνει, πλέον, τιμές, σε διάστημα διαφορετικού μήκους.
- $h(A\mathbf{X}) = h(\mathbf{X}) + \log|\det(A)|$, όπου $\det(A)$ η ορίζουσα του A.