ΕΕ728

Προχωρημένα Θέματα Θεωρίας Πληροφορίας

Δημήτρης - Αλέξανδρος Τουμπακάρης

8ο Μάθημα – 6 Μαΐου 2009

Ανακεφαλαίωση προηγούμενου μαθήματος

- Θεώρημα Κωδικοποίησης Καναλιού για διακριτά κανάλια χωρίς μνήμη – Απόδειξη ευθέος.

Προεπισκόπηση σημερινού μαθήματος

- Απόδειξη θεωρήματος Κωδικοποίησης Καναλιού για διακριτά κανάλια χωρίς μνήμη (αντίστροφο).
- Παρατηρήσεις και Θεωρήματα για τη Χωρητικότητα διακριτών καναλιών χωρίς μνήμη.
- Χωρητικότητα καναλιών με ανάδραση.
- Αποκωδικοποίηση Μέγιστης Πιθανοφάνειας και Εκθέτης Σφάλματος
- Θεώρημα Διαχωρισμού Πηγής - Καναλιού

$I(X^n;Y^n) \leq nC$

Θα αποδείξουμε, κατ΄ αρχήν, ότι, για Διακριτά Κανάλια Χωρίς Μνήμη, η πληροφοριακή χωρητικότητα ανά χρήση του καναλιού δεν αυξάνει εάν το κανάλι χρησιμοποιηθεί πολλές φορές. Δηλαδή, $I(X^n;Y^n) \leq nC$ για οποιαδήποτε $p(x)$, όπου $C = \max_{p(x)} I(X;Y)$.

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

$$\stackrel{(a)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \stackrel{(b)}{\leq} \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i)$$

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

(a) Το κανάλι δεν έχει μνήμη, επομένως η έξοδος τη χρονική στιγμή i εξαρτάται μόνο από την είσοδο τη χρονική στιγμή i. Επίσης, δε χρησιμοποιείται ανάδραση. (b) Η από κοινού εντροπία δεν υπερβαίνει το άθροισμα των εντροπιών.

Ανισότητα **Fano**

- Για την απόδειξη του αντιστρόφου του Θεωρήματος Κωδικοποίησης Καναλιού θα χρησιμοποιήσουμε την Ανισότητα **Fano**.

- Είδαμε ότι, για κάθε εκτιμητή $\hat{X} = g(Y)$,

  $$H(X|Y) \leq H(X|\hat{X}) \leq H(P_e) + P_e \log|\mathcal{X}| \Rightarrow H(X|\hat{X}) \leq 1 + P_e \log|\mathcal{X}|,$$

  όπου $P_e = \Pr\{\hat{X} \neq X\}$.

- Εάν θεωρήσουμε Διακριτό Κανάλι Χωρίς Μνήμη με βιβλίο κωδίκων $\mathcal{C}$ και ομοιόμορφα κατανεμημένα μηνύματα W,

  $$H(W|\hat{W}) \leq 1 + P_e^{(n)} nR, \text{ όπου } P_e^{(n)} = \Pr\{W \neq \hat{W}\}.$$

Θεώρημα Κωδικοποίησης Καναλιού – Απόδειξη αντιστρόφου

- Θα δείξουμε ότι, για κάθε κώδικα $(2^{nR}, n)$ με $\lambda^{(n)} \to 0$, πρέπει να ισχύει $R \leq C$. Δεδομένου ότι $\lambda^{(n)} \to 0$ και η μέση πιθανότητα σφάλματος $P_e^{(n)} \to 0$.

- Έστω ότι ο δέκτης αποφασίζει ποια ακολουθία μεταδόθηκε με βάση κάποια συνάρτηση αποκωδικοποίησης $\hat{W} = g(Y^n)$. Ισχύει $W \to X^n(W) \to Y^n \to \hat{W}$.

- Έστω, επίσης, ότι το μήνυμα που στέλνεται στο κανάλι επιλέγεται με βάση ομοιόμορφη κατανομή στο σύνολο των πιθανών μηνυμάτων $\{1, 2, \ldots, 2^{nR}\}$. Επομένως, $\Pr\{\hat{W} \neq W\} = P_e^{(n)} = \frac{1}{2^{nR}} \sum_i \lambda_i$.

Θεώρημα Κωδικοποίησης Καναλιού – Απόδειξη αντιστρόφου (2)

- Συνεπώς,

$$
\begin{aligned}
nR &\overset{(a)}{=} H(W) \overset{(b)}{=} H(W|\hat{W}) + I(W;\hat{W}) \overset{(c)}{\leq} 1 + P_e^{(n)} nR + I(W;\hat{W}) \\
&\overset{(d)}{\leq} 1 + P_e^{(n)} nR + I(X^n;Y^n) \overset{(e)}{\leq} 1 + P_e^{(n)} nR + nC.
\end{aligned}
$$

(a) W ομοιόμορφη τ.μ., (b) σχέση αμοιβαίας πληροφορίας – εντροπίας, (c) ανισότητα Fano, (d) ανισότητα επεξεργασίας δεδομένων, (e) $I(X^n;Y^n) \leq nC$.

Θεώρημα Κωδικοποίησης Καναλιού
Απόδειξη αντιστρόφου (3)

$$nR \leq 1 + P_e^{(n)} nR + nC \Rightarrow R \leq P_e^{(n)} R + \frac{1}{n} + C.$$

- Από την υπόθεση ότι $\lambda^{(n)} \to 0$, $P_e^{(n)} R \to 0$ για $n \to \infty$. Επομένως, για $n \to \infty$,
$$R \leq C.$$

- Λύνοντας ως προς $P_e^{(n)}$, $P_e^{(n)} \geq 1 - \frac{C}{R} - \frac{1}{nR}$. Συνεπώς, εάν $R > C$, $P_e^{(n)} > 0$ για $n \to \infty$.

Θεώρημα Κωδικοποίησης Καναλιού
Απόδειξη αντιστρόφου (4)

- Το αποτέλεσμα αυτό ονομάζεται ασθενές αντίστροφο του Θεωρήματος Κωδικοποίησης Καναλιού. Αποδεικνύεται (ισχυρό αντίστροφο) ότι, εάν $R > C$, $P_e^{(n)} \to 1$ εκθετικά.

- Συνεπώς, η χωρητικότητα καναλιού C αποτελεί μια πολύ σαφή διαχωριστική γραμμή: Όταν $R < C$ η πιθανότητα σφάλματος τείνει εκθετικά στο 0. Αντίθετα, όταν $R > C$, η πιθανότητα σφάλματος τείνει εκθετικά στο 1.

Παρατηρήσεις και θεωρήματα σχετικά με τη χωρητικότητα

- Απόδειξη Θεωρήματος Κωδικοποίησης (αντίστροφο)
- Παρατηρήσεις και θεωρήματα σχετικά με τη χωρητικότητα
- Χωρητικότητα καναλιών με ανάδραση
- Αποκωδικοποίηση Μέγιστης Πιθανοφάνειας και Εκθέτης Σφάλματος
- Θεώρημα Διαχωρισμού Πηγής - Καναλιού

Μεγιστοποίηση κοίλης συνάρτησης κατανομής πιθανότητας

- Θεωρούμε συνάρτηση $f(\mathbf{p}) : \mathbf{R}^n \to \mathbf{R}$ η οποία είναι κοίλη $\cap$ ως προς $\mathbf{p}$.

- Έστω, επίσης, ότι το $\mathbf{p}$ είναι κατανομή (διάνυσμα πιθανότητας), δηλαδή $p_i \geq 0,\ i = 1, \ldots, n$ και $\sum_{i=1}^{n} p_i = \mathbf{1}^T \mathbf{p} = 1$.

- Τέλος, θεωρούμε ότι οι μερικές παράγωγοι $\partial f(\mathbf{p})/\partial p_i$ ορίζονται και ότι είναι συνεχείς με μοναδική εξαίρεση το $\lim_{p_i \to 0} \partial f(\mathbf{p})/\partial p_i$ που μπορεί να είναι και ∞.

Μεγιστοποίηση κοίλης συνάρτησης κατανομής πιθανότητας (συνέχεια)

- Αποδεικνύεται ότι οι παρακάτω συνθήκες είναι ικανές και αναγκαίες για να μεγιστοποιείται η $f()$ στο σημείο (κατανομή) $\mathbf{p}$.

$$\frac{\partial f(\mathbf{p})}{\partial p_i} = \lambda, \text{ για όλα τα } i \text{ για τα οποία } p_i > 0$$
$$\frac{\partial f(\mathbf{p})}{\partial p_i} \leq \lambda, \text{ για όλα τα } i \text{ για τα οποία } p_i = 0$$

  για κάποια τιμή της παραμέτρου λ.

- Για την απόδειξη δείτε π.χ. Gallager Theorem 4.4.1.

Μεγιστοποίηση αμοιβαίας πληροφορίας

- Με χρήση του προηγούμενου θεωρήματος και του ότι η $I(X;Y)$ είναι κοίλη $\cap$ συνάρτηση της κατανομής εισόδου $p(x)$ για δεδομένο κανάλι $p(y|x)$, αποδεικνύεται ότι οι παρακάτω δύο συνθήκες αποτελούν ικανή και αναγκαία συνθήκη για να επιτυγχάνει μια κατανομή $\mathbf{p}^*$ τη χωρητικότητα.

$$I(X = x_i; Y) = C, \text{ για όλα τα } i \text{ για τα οποία } p^*_{x_i} > 0$$
$$I(X = x_i; Y) \leq C, \text{ για όλα τα } i \text{ για τα οποία } p^*_{x_i} = 0$$

όπου $I(X = x_i; Y) = \sum_{y \in |\mathcal{Y}|} p(y|x_i) \log \frac{p(y|x_i)}{p(y)}$ η αμοιβαία πληροφορία μεταξύ $X = x_i$ και Y.

Μεγιστοποίηση αμοιβαίας πληροφορίας (συνέχεια)

- Το αποτέλεσμα αυτό έχει μια διαισθητική επεξήγηση: Εάν για $x_i \neq x_j$ $I(X = x_i; Y) > I(X = x_j; Y)$, μπορούμε να αυξήσουμε την $I(X;Y) = \sum_{x_k} p(x_k) I(X = x_k; Y)$ χρησιμοποιώντας τη x_i πιο συχνά και τη x_j λιγότερο συχνά (αλλάζοντας τις $p(x_i)$ και $p(x_j)$).

- Αυτό έχει ως αποτέλεσμα να αλλάξει η $p(y) = \sum_{x_k} p(x_k) p(y|x_k)$.

- Τελικά, η διαδικασία αυτή θα ισορροπήσει σε σημείο όπου όλες οι $I(X = x_i; Y)$ εκτός, ίσως, από κάποιες που αντιστοιχούν σε κακές εισόδους, θα ισούνται μεταξύ τους (και, επομένως, και με τη χωρητικότητα, C).

Άλλες ενδιαφέρουσες ιδιότητες και αποτελέσματα

- Αναφέρουμε, τέλος, 3 ενδιαφέροντα πορίσματα. Για αποδείξεις δείτε π.χ. Gallager Κεφ. 4.5.

- Πόρισμα 1: Για οποιαδήποτε κατανομή εισόδου, $p^*(x)$, που επιτυγχάνει τη χωρητικότητα σε διακριτό κανάλι χωρίς μνήμη, όλες οι πιθανότητες συμβόλων εξόδου, $p(y)$, είναι αυστηρώς θετικές (αρκεί για κάθε έξοδο να υπάρχει τουλάχιστον μία είσοδος που οδηγεί σε αυτήν).

- Πόρισμα 2: Η κατανομή εξόδου, $p^*(y)$, για την οποία $I(X;Y) = C$ είναι μοναδική. Όλες οι κατανομές εισόδου, $p(x)$, για τις οποίες $\sum_{x\in|\mathcal{X}|} p(x)p(y|x) = p^*(y)$ επιτυγχάνουν τη χωρητικότητα.

Άλλες ενδιαφέρουσες ιδιότητες και αποτελέσματα (συνέχεια)

- Πόρισμα 3: Έστω m ο ελάχιστος αριθμός συμβόλων εισόδου που μπορούν να χρησιμοποιηθούν (με μη μηδενική πιθανότητα) για να επιτευχθεί μετάδοση με τη χωρητικότητα. Έστω $\mathcal{A}$ ένα τέτοιο σύνολο m συμβόλων εισόδου. Ισχύει $m \leq |\mathcal{Y}|$. Επίσης, η κατανομή $p(x)$ στα στοιχεία του $\mathcal{A}$ που επιτυγχάνει τη χωρητικότητα είναι μοναδική.

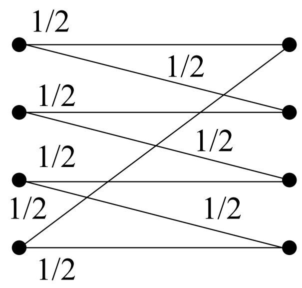
Πώς υπολογίζουμε τη χωρητικότητα;

- Γενικά, ο υπολογισμός της χωρητικότητας δεν είναι εύκολη υπόθεση.
- Σε μερικές, ειδικές, περιπτώσεις μπορούμε να χρησιμοποιήσουμε ιδιότητες όπως, π.χ. στην περίπτωση συμμετρικών καναλιών.
- Άλλες φορές μπορούμε να "μαντέψουμε" την κατανομή εισόδου και να δείξουμε ότι επιτυγχάνει ένα άνω φράγμα για τη χωρητικότητα (όπως κάναμε για το συμμετρικό κανάλι).
- Στη γενική περίπτωση καταφεύγουμε σε αριθμητικές μεθόδους με χρήση υπολογιστή. Μια ευρέως χρησιμοποιούμενη μέθοδος είναι των Blahut & Arimoto. Τα τελευταία χρόνια έχουν προταθεί βελτιώσεις που συγκλίνουν πολύ πιο γρήγορα σε σχέση με τον αρχικό αλγόριθμο.

Χωρητικότητα καναλιών με ανάδραση (**feedback**)

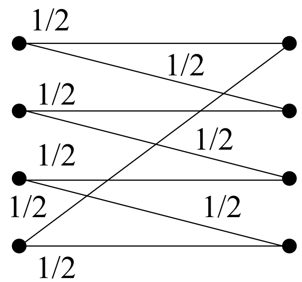
- Απόδειξη Θεωρήματος Κωδικοποίησης (αντίστροφο)
- Παρατηρήσεις και θεωρήματα σχετικά με τη χωρητικότητα
- Χωρητικότητα καναλιών με ανάδραση
- Αποκωδικοποίηση Μέγιστης Πιθανοφάνειας και Εκθέτης Σφάλματος
- Θεώρημα Διαχωρισμού Πηγής - Καναλιού

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 $= C$.

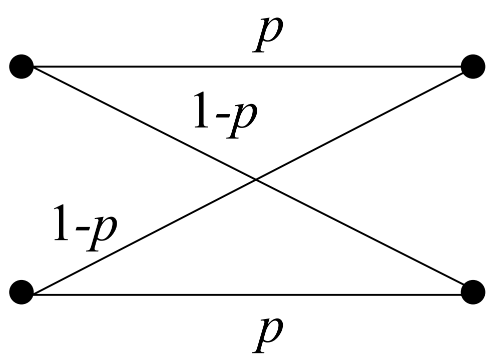
Παράδειγμα 8.1 (συνέχεια)



- 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;

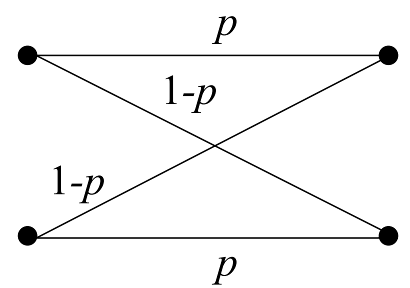
Παράδειγμα 8.2



- Ας θεωρήσουμε, τώρα, το δυαδικό συμμετρικό κανάλι.

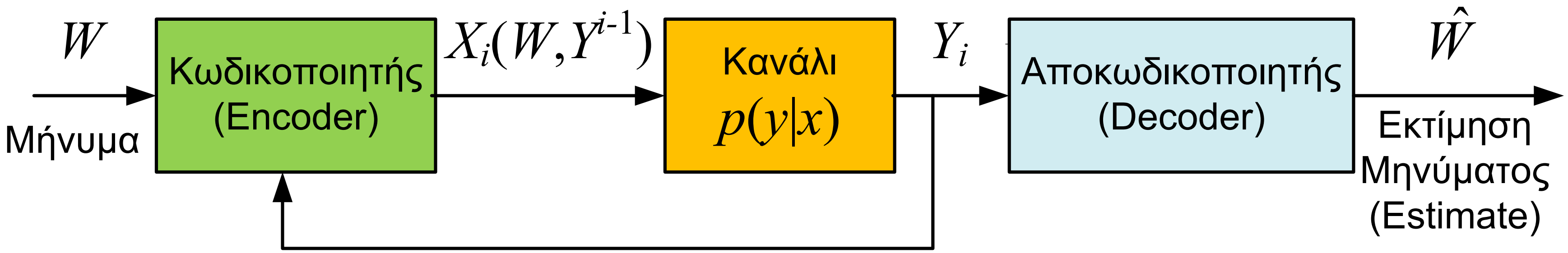
- Γνωρίζουμε ότι $C = 1 - H(p)$ και ότι η χωρητικότητα επιτυγχάνεται χρησιμοποιώντας και τα δύο μηνύματα με ίση πιθανότητα. Επομένως, κάθε φορά που στέλνουμε ένα από τα δύο μηνύματα στο κανάλι δε γνωρίζουμε εάν το μήνυμα μεταδόθηκε επιτυχώς. Η πιθανότητα σφάλματος ανά μετάδοση είναι μη μηδενική.

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 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

- 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 $X^n = X^n(W)$.

- 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΄ αρχήν, 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 W είναι ομοιόμορφα κατανεμημένο στο σύνολο $\{1, 2, \ldots, 2^{nR}\}$.

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

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

Χωρητικότητα καναλιού με ανάδραση (3)

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

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

$$\begin{aligned} I(W; Y^n) &= H(Y^n) - H(Y^n|W) \overset{(a)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, Y_2, \ldots, Y_{i-1}, W) \\ &\overset{(b)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, Y_2, \ldots, Y_{i-1}, W, X_i) \overset{(c)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &\le \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) = \sum_{i=1}^{n} I(X_i; Y_i) \le nC. \end{aligned}$$

(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, οπότε η έξοδος τη χρονική στιγμή i εξαρτάται μόνο από την είσοδο X_i.

Χωρητικότητα καναλιού με ανάδραση (4)

- Επομένως, $I(W;Y^n) \leq nC$, και

$$nR \leq P_e^{(n)} nR + 1 + nC.$$

- 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

- Απόδειξη Θεωρήματος Κωδικοποίησης (αντίστροφο)
- Παρατηρήσεις και θεωρήματα σχετικά με τη χωρητικότητα
- Χωρητικότητα καναλιών με ανάδραση
- **Αποκωδικοποίηση Μέγιστης Πιθανοφάνειας και Εκθέτης Σφάλματος**
- Θεώρημα Διαχωρισμού Πηγής - Καναλιού

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 δε 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 detection).

- 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: $w = g(y^n)$, τέτοιο ώστε

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

Εναλλακτική έκφραση:

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

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 γραφεί ως:

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

- 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:
  - Σε κάποια συστήματα, τα μηνύματα που στέλνονται είναι ισοπίθανα, 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 χρησιμοποιειται 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

- Απόδειξη Θεωρήματος Κωδικοποίησης (αντίστροφο)
- Παρατηρήσεις και θεωρήματα σχετικά με τη χωρητικότητα
- Χωρητικότητα καναλιών με ανάδραση
- Αποκωδικοποίηση Μέγιστης Πιθανοφάνειας και Εκθέτης Σφάλματος
- 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)

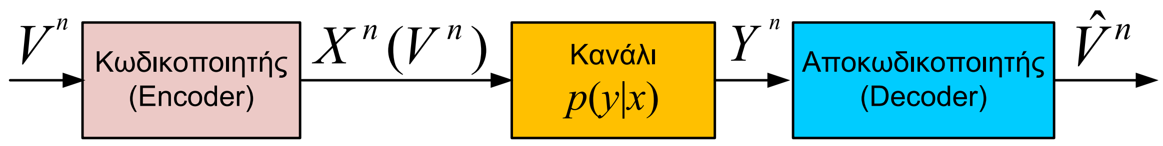
- Ειδικότερα, είναι βέλτιστο να συμπιέσουμε την πηγή κοντά στο ρυθμό εντροπίας της και μετά να μεταδώσουμε τη συμπιεσμένη ακολουθία στο κανάλι ή μήπως υπάρχει πιο αποδοτικός τρόπος μετάδοσης (και, άρα, τρόπος να μεταδώσουμε με μεγαλύτερο ρυθμο;)

- Θα αποδείξουμε ότι η μετάδοση με συμπίεση της πηγής και, στη συνέχεια, 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, εφόσον η πληροφορία μιας πηγής μεταδίδεται με αυθαίρετα μικρή πιθανότητα σφάλματος στο κανάλι, ισχύει πάντα $H(\mathcal{X}) < C$.

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 τη 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 <u>στάσιμη</u> και <u>εργοδική</u>.

- 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

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

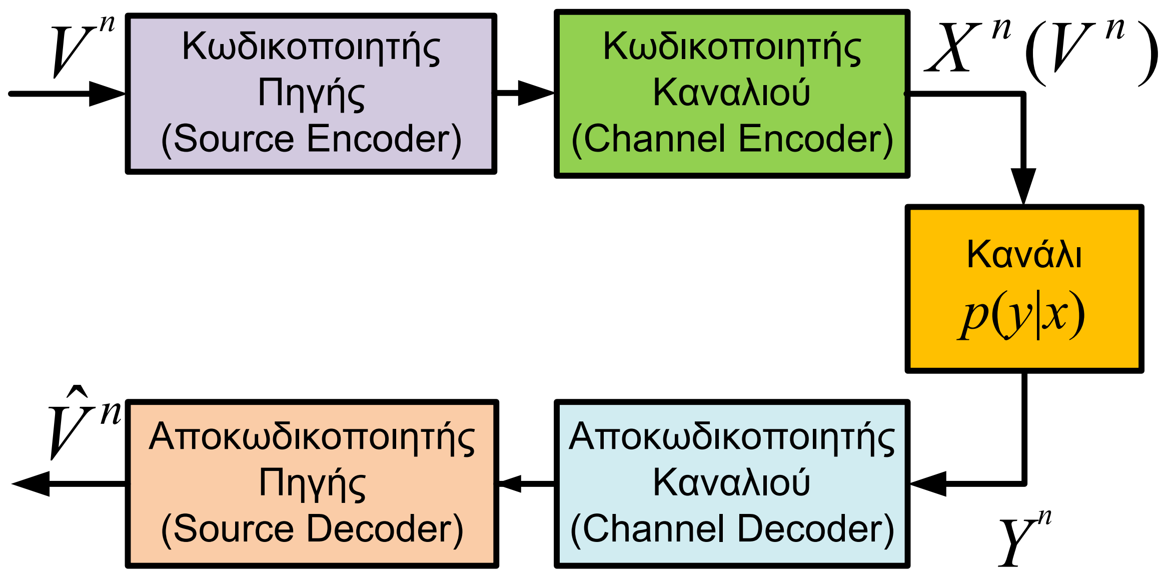
  όπου I η συνάρτηση-δείκτης και $g(\cdot)$ η συνάρτηση αποκωδικοποίησης.

- <u>Θεώρημα Διαχωρισμού Πηγής - Καναλιού</u> (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 $\Pr\{\hat{V}^n \neq V^n\} \rightarrow 0$.

- Αντίστροφα, για κάθε στάσιμη και εργοδική στοχαστική ανέλιξη, εάν $H(\mathcal{V}) > C$, η πιθανότητα σφάλματος δε 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 $H(\mathcal{V})$

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

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

Θεώρημα Διαχωρισμού Πηγής - Καναλιού
Απόδειξη αντιστρόφου (συνέχεια)

$$
\begin{aligned}
H(\mathcal{V}) &\le \frac{1}{n}\Big(1 + n\,\Pr\{\hat{V}^n \ne V^n\}\log|\mathcal{V}|\Big) + \frac{1}{n} I(V^n;\hat{V}^n) \\
&\overset{(a)}{\le} \frac{1}{n}\Big(1 + n\,\Pr\{\hat{V}^n \ne V^n\}\log|\mathcal{V}|\Big) + \frac{1}{n} I(X^n;Y^n) \\
&\overset{(b)}{\le} \frac{1}{n} + \Pr\{\hat{V}^n \ne V^n\}\log|\mathcal{V}| + C.
\end{aligned}
$$

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

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

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

Ανακεφαλαίωση μαθήματος

- Θεώρημα Κωδικοποίησης Καναλιού (για Διακριτά Κανάλια Χωρίς Μνήμη)
  - Αντίστροφο: Με χρήση της ανισότητας Fano δείξαμε ότι δεν υπάρχει κώδικας με $P_e^{(n)} \to 0$ που να επιτυγχάνει $R > C$.
- Χωρητικότητα Διακριτών Καναλιών Χωρίς Μνήμη με Ανάδραση.
  - Η χωρητικότητα διακριτού καναλιού χωρίς μνήμη δεν αυξάνει εάν χρησιμοποιήσουμε ανάδραση!
- Θεώρημα διαχωρισμού πηγής-καναλιού
  - Σε διακριτά κανάλια χωρίς μνήμη η κωδικοποίηση πηγής και η κωδικοποίηση καναλιού μπορούν να γίνουν ανεξάρτητα, χωρίς να επηρεαστεί το μέγιστο ποσό πληροφορίας της πηγής που μπορούμε να μεταδώσουμε με χρήση του καναλιού.

Προεπισκόπηση επόμενου μαθήματος

- Συνεχείς τ.μ. και κανάλια διακριτού χρόνου αλλά συνεχών τιμών.
- Ποσότητες Θεωρίας Πληροφορίας για συνεχείς τ.μ.: Διαφορική Εντροπία, Σχετική Εντροπία και Αμοιβαία Πληροφορία για συνεχείς τ.μ.
- Ιδιότητες Διαφορικής Εντροπίας.
- AEP για συνεχείς τ.μ.
- Η (πολυμεταβλητή) Γκαουσιανή κατανομή και η εντροπία της.
- Το Γκαουσιανό κανάλι. Χωρητικότητα και Θεώρημα Κωδικοποίησης.