ΕΕ728 Προχωρημένα Θέματα Θεωρίας Πληροφορίας 8η διάλεξη (2η έκδοση, 16/4/2013)

Δημήτρης-Αλέξανδρος Τουμπακάρης

Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών

2 Απριλίου 2013

Περιεχόμενα 8ης διάλεξης

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 $+\infty$.

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.

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

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

- **Πόρισμα 8.2.** Η κατανομή εξόδου, $p^*(y)$, για την οποία $I(X;Y) = C$ είναι <u>μοναδική</u>. Όλες οι κατανομές εισόδου, $p(x)$, για τις οποίες $\sum_{x\in|\mathcal{X}|} p(x)p(y|x) = p^*(y)$ επιτυγχάνουν τη χωρητικότητα.

- **Πόρισμα 8.3.** 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.

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 *ασυμπτωτικώς 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 $p(y^n|x^n(m))$ να ληφθεί η ακολουθία y^n στο δέκτη δεδομένου ότι εστάλη ακολουθία $x^n(m)$ η οποία αντιστοιχεί στο μήνυμα m (η κωδική λέξη του μηνύματος m).
- Από το Θεώρημα Ολικής πιθανότητας, $\Pr\{\hat{m} = m\}(= 1 - \Pr\{\hat{m} \neq m\}) = \sum_{y^n=1}^{|\mathcal{Y}|^n} p(y^n) \Pr\{\hat{m} = m|y^n\} = \sum_{y^n=1}^{|\mathcal{Y}|^n} p(y^n) \Pr\{\hat{x}^n = x^n|y^n\}$.
- Επειδή $p(y^n) \geq 0$, για να μεγιστοποιήσουμε την $\Pr\{\hat{m} = m\}$ αρκεί να μεγιστοποιήσουμε κάθε όρο $\Pr\{\hat{x}^n = x^n|y^n\}$.

Αποκωδικοποίηση Μέγιστης εκ των Υστέρων Πιθανότητας (3)

- Από τον κανόνα του Bayes,

$$p(m|y^n) = \frac{p(y^n|x^n(m))p(m)}{p(y^n)},$$

$$\text{όπου } p(y^n) = \sum_{m=1}^{|\mathcal{M}|} p(m)p(y^n|x^n(m)).$$

- Για να ελαχιστοποιηθεί η πιθανότητα σφάλματος, πρέπει να επιλεγεί το μήνυμα m το οποίο μεγιστοποιεί την εκ των υστέρων (a posteriori) πιθανότητα του m δεδομένης της ληφθείσας ακολουθίας y^n ($p(m|y^n)$).

Κανόνας αποκωδικοποίησης MAP

Κανόνας αποκωδικοποίησης MAP

$\hat{m} = g(y^n)$, τέτοιο ώστε

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

Εναλλακτική έκφραση

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

Αποκωδικοποίηση Μέγιστης εκ των Υστέρων Πιθανότητας (4)

- Με χρήση του κανόνα του Bayes,

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

- Επομένως, ο κανόνας MAP μπορεί να γραφτεί ως:

Κανόνας MAP

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

- Για κανάλι χωρίς μνήμη,

Κανόνας MAP για κανάλι χωρίς μνήμη

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

Αποκωδικοποίηση Μέγιστης Πιθανοφάνειας (Maximum Likelihood (ML) decoding)

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

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

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

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

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, αρκεί ο ρυθμός μετάδοσης να μην υπερβαίνει τη χωρητικότητα.
- Αντιστρόφως, 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 Ισο-διαμέρισης. 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 8.4. (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 $P_{e,m}^{(n)}$ που αντιστοιχεί στην κωδική λέξη $x^n(m)$ θα είναι $\leq \exp\{-nE_r(R)\}$. Ωστόσο, αποδεικνύεται (Gallager 5.6 Corollary 2) ότι υπάρχει κώδικας $(2^{nR}, n)$ τέτοιος ώστε $P_{e,m}^{(n)} \leq 4\exp\{-nE_r(R)\}$ για όλα τα $m = 1, 2, \ldots, 2^{nR}$.

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 ένα από 2^{nR} μηνύματα/χρήση 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;

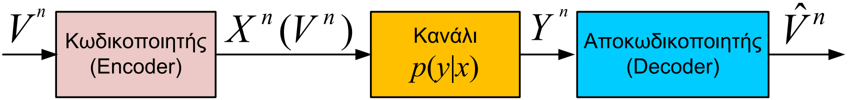
Θεώρημα Διαχωρισμού Πηγής - Καναλιού – Εισαγωγή (2)

- 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, εάν $H(\mathcal{X}) < C$, μπορούμε να συμπιέσουμε την πηγή και να μεταδώσουμε την πληροφορία που παράγει μέσω του καναλιού. Αντιστρόφως, προκειμένου να είναι εφικτό η πληροφορία μιας πηγής να μεταδοθεί 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 ακολουθίας της πηγής με βάση τη ληφθείσα ακολουθία Y^n. 'Οταν $\hat{V}^n \neq V^n$ εμφανίζεται σφάλμα στο δέκτη.

Θεώρημα Διαχωρισμού Πηγής - Καναλιού (συνέχεια)

- Η πιθανότητα σφάλματος ισούται με

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

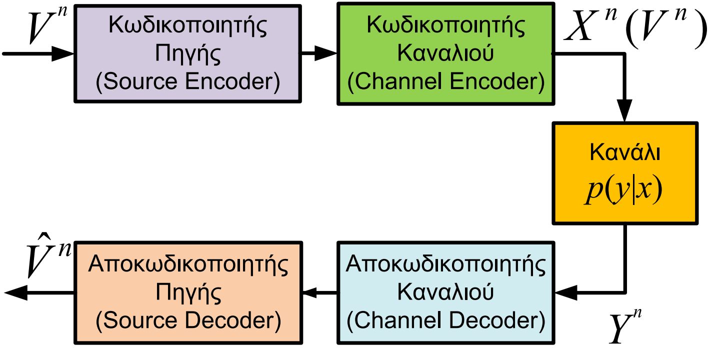
όπου I η συνάρτηση-δείκτης και $g(\cdot)$ η συνάρτηση αποκωδικοποίησης.

- Θεώρημα 8.5. (Διαχωρισμού Πηγής - Καναλιού – ευθύ): Έστω $V_1, V_2, \ldots, V_n$ στοχαστική διαδικασία με πεπερασμένο αλφάβητο η οποία ικανοποιεί το AEP, και για την οποία ισχύει $H(\mathcal{V}) < C$, όπου C είναι η χωρητικότητα του διακριτού καναλιού χωρίς μνήμη μέσω του οποίου γίνεται η μετάδοση. Υπάρχει κώδικας πηγής-καναλιού με πιθανότητα σφάλματος $\Pr\{\hat{V}^n \neq V^n\} \to 0$.

- Αντιστρόφως, για κάθε στάσιμη και εργοδική στοχαστική διαδικασία, εάν $H(\mathcal{V}) > C$, η πιθανότητα σφάλματος δεν μπορεί να περιοριστεί αυθαίρετα κοντά στο 0 και, εμπομένως, δεν είναι δυνατή η μετάδοση της στοχαστικής διαδικασίας 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. Σε κάθε τυπική ακολουθία αντιστοιχίζουμε μία κωδική λέξη από το βιβλίο κωδίκων. Επομένως, χρειαζόμαστε το πολύ $2^{n(H(\mathcal{V})+\epsilon)}$ κωδικές λέξεις.
- Προκειμένου να μεταδώσουμε μία από $2^{n(H(\mathcal{V})+\epsilon)}$ κωδικές λέξεις 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

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

Θεώρημα Διαχωρισμού Πηγής - Καναλιού Απόδειξη ευθέος (3)

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

- Για αρκούντως μεγάλο n, από το AEP, $\Pr\left\{V^n \notin A_\epsilon^{(n)}\right\} \leq \epsilon$.
- Ομοίως, από το Joint AEP, για αρκούντως μεγάλο n, και δεδομένου ότι $H(\mathcal{V}) + \epsilon = R < C$, $\Pr\left\{g(Y^n) \neq V^n | V^n \in A_\epsilon^{(n)}\right\} \leq \epsilon$.
- 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 βημάτων (συμπίεση και κωδικοποίηση καναλιού), μπορούμε να μεταδώσουμε με αυθαίρετα μικρή πιθανότητα σφάλματος εάν $H(\mathcal{V}) < C$.