EE728

Προχωρημένα Θέματα Θεωρίας Πληροφορίας

Δημήτρης - Αλέξανδρος Τουμπακάρης

8ο Μάθημα – 9 Μαΐου 2008

Ανακεφαλαίωση προηγούμενου μαθήματος

- Αποδείξαμε το Θεώρημα Κωδικοποίησης Καναλιού (για Διακριτά Κανάλια Χωρίς Μνήμη)
  - Ευθύ: Αποδείξαμε ότι, για δέκτη που χρησιμοποιεί αποκωδικοποίηση με βάση την Από Κοινού Τυπικότητα, για $n \to \infty$, η πιθανότητα να αποκωδικοποιήσουμε λάθος σύμβολο ή να μη μπορούμε να αποκωδικοποιήσουμε τείνει στο 0, εφόσον $R < C = \max_{p(x)} I(X;Y)$.
  - Μπορεί να αποδειχθεί και για αποκωδικοποιητή ML.
  - Αντίστροφο: Με χρήση της ανισότητας Fano δείξαμε ότι δεν υπάρχει κώδικας με $P_e^{(n)} \to 0$ που να επιτυγχάνει $R > C$.
- Η χρήση ανάδρασης (feedback) δεν αυξάνει τη χωρητικότητα διακριτού καναλιού χωρίς μνήμη!
- Ο βέλτιστος αποκωδικοποιητής βασίζεται σε ανίχνευση μέγιστης πιθανοφάνειας (ML). Ένα άνω φράγμα για την πιθανότητα σφάλματος για δεδομένο μήκος κώδικα n μπορεί να προσδιοριστεί με χρήση του Εκθέτη Σφάλματος.

Προεπισκόπηση σημερινού μαθήματος

- Θα αποδείξουμε το Θεώρημα Διαχωρισμού Πηγής - Καναλιού.
- Θα αρχίσουμε να εξετάζουμε την περίπτωση συνεχών τ.μ. και τα αντίστοιχα μεγέθη της Θεωρίας Πληροφορίας.

Περιεχόμενα σημερινού μαθήματος

- Θεώρημα Διαχωρισμού Πηγής - Καναλιού
- Συνεχείς τ.μ. και Διαφορική Εντροπία.
- Ποσότητες Θεωρίας Πληροφορίας για συνεχείς τ.μ. Ιδιότητες.

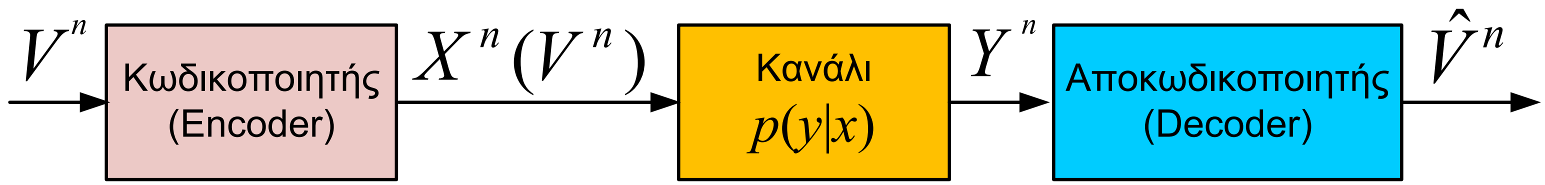
Θεώρημα Διαχωρισμού Πηγής - Καναλιού – Εισαγωγή

- 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;
- 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 (συνέχεια)

- 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

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

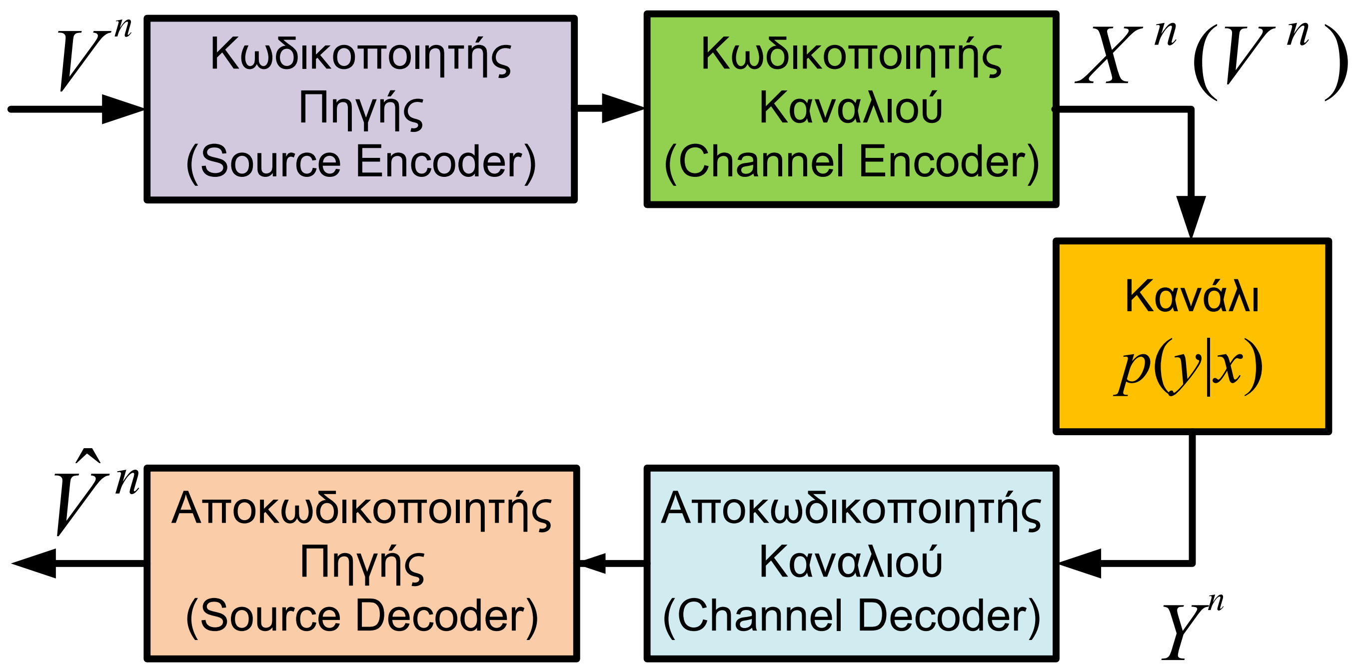
  όπου I η συνάρτηση-δείκτης και $g(\cdot)$ η συνάρτηση αποκωδικοποίησης.

- 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 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,

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

Συνεχείς τ.μ. και Διαφορική Εντροπία

- Θεώρημα Διαχωρισμού Πηγής - Καναλιού
- **Συνεχείς τ.μ. και Διαφορική Εντροπία**
- Ποσότητες Θεωρίας Πληροφορίας για συνεχείς τ.μ. Ιδιότητες.

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.
- 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.

Παράδειγμα 8.1. – 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.

Παράδειγμα 8.2. – Εντροπία Γκαουσιανής τ.μ.

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

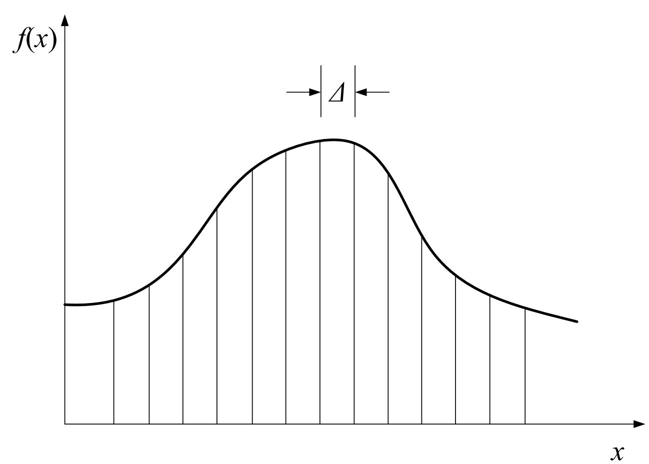
$$X \sim \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{-x^2}{2\sigma^2}}.$$

- Για τη διαφορική εντροπία ισχύει

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

Παράδειγμα 8.3. – Εντροπία διακριτής αναπαράστασης συνεχούς τ.μ.

- Έστω συνεχής τ.μ. X με συνάρτηση πυκνότητας πιθανότητας $f(x)$. Ξωρίζουμε 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:

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

Παράδειγμα 8.3. – 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 X^Δ ισχύει

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

Παράδειγμα 8.3. – 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.
- 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, $H(X^{\Delta}) \approx h(X) + n$.

Ποσότητες Θεωρίας Πληροφορίας για συνεχείς τ.μ. Ιδιότητες

- Θεώρημα Διαχωρισμού Πηγής - Καναλιού
- Συνεχείς τ.μ. και Διαφορική Εντροπία
- Ποσότητες Θεωρίας Πληροφορίας για συνεχείς τ.μ. Ιδιότητες

Από κοινού και υπό συνθήκη Διαφορική Εντροπία

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

- Από κοινού διαφορική εντροπία: $h(X_1, X_2, \ldots, X_n) = -\int f(x^n) \log f(x^n) dx^n$, όπου $f(x^n) = f(x_1, x_2, \ldots, x_n)$.
- Υπό συνθήκη διαφορική εντροπία: $h(X|Y) = -\int f(x, y) \log f(x|y) dx dy$.
- Όπως και στην περίπτωση διακριτών τ.μ., εάν όλες οι ποσότητες είναι πεπερασμένες,

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

Παράδειγμα 8.4 – 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., $I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X) = h(X) + h(Y) - h(X,Y)$.
- Εάν δεν ορίζεται $f(x,y)$, μπορεί να χρησιμοποιηθεί ο πλέον γενικός ορισμός της Αμοιβαίας Πληροφορίας

$$I(X;Y) = \sum_{\mathcal{P},\mathcal{Q}} I([X]_{\mathcal{P}};[Y]_{\mathcal{Q}}),$$

όπου $\mathcal{P}$ και $\mathcal{Q}$ αποτελούν πεπερασμένες διαμερίσεις (partitions) των $\mathcal{X}$ και $\mathcal{Y}$, αντίστοιχα (περισσότερες λεπτομέρειες στο βιβλίο του Cover).

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

- $D(f||g) \geq 0$, με ισότητα όταν $f = g$. 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;
- $h(X|Y) \leq h(X)$ 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 κατ' ευθείαν από τον ορισμό.
  - Η διαφορική εντροπία είναι αναλλοίωτη σε μετάθεση.
- $h(aX) = h(X) + \log|a|$. Για την απόδειξη δείτε π.χ. Cover Theorem 8.6.4.
- $h(A\mathbf{X}) = h(\mathbf{X}) + \log|\det(A)|$, όπου $\det(A)$ η ορίζουσα του A.

Μεγιστοποίηση Διαφορικής Εντροπίας

- Έστω τυχαίο διάνυσμα $\mathbf{X} \in \mathbf{R}^n$ με μέση τιμή $\mathbf{m} = 0$ και πίνακα συσχέτισης $K = E\mathbf{X}\mathbf{X}^T$ (δηλαδή $K_{ij} = EX_iX_j,\ 1 \le i, j \le n$).
  $h(\mathbf{X}) \le \frac{1}{2}\log(2\pi e)^n|K|$, με $=$ εάν και μόνο εάν $\mathbf{X} \sim \mathcal{N}(0, K)$.
- Επομένως, για δεδομένο πίνακα συσχέτισης, η συνεχής κατανομή που μεγιστοποιεί την εντροπία είναι η Γκαουσιανή!
- Θα μας χρησιμεύσει στον υπολογισμό της χωρητικότητας του Γκαουσιανού Καναλιού.
- Για βαθμωτή συνεχή τ.μ. X με μέση τιμή $m = 0$ και μεταβλητότητα σ^2, η κατανομή που μεγιστοποιεί την $h(X)$ είναι η $\mathcal{N}(0, \sigma^2)$.
- Δεδομένου ότι $h(X + c) = h(X)$, μπορούμε να επεκτείνουμε το αποτέλεσμα και σε τ.μ. με μη μηδενική μέση τιμή.
- Μεταξύ συνεχών τ.μ. με την ίδια ισχύ, οι πιο "αβέβαιες" είναι αυτές που ακολουθούν Γκαουσιανή κατανομή.

Μεγιστοποίηση Διαφορικής Εντροπίας – Απόδειξη

- Έστω $g(\mathbf{x})$ οποιαδήποτε συνάρτηση πυκνότητας πιθανότητας που ικανοποιεί τον περιορισμό συσχέτισης $\int g(\mathbf{x}) x_i x_j d\mathbf{x} = K_{ij}$ για όλα τα i, j. Έστω, επίσης ϕ_K η συνάρτηση πυκνότητας πιθανότητας διανύσματος που ακολουθεί Γκαουσιανή κατανομή $\mathcal{N}(0, K)$:

$$\phi_K(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}\mathbf{x}^T K^{-1}\mathbf{x}} \Rightarrow \log \phi_K(\mathbf{x}) = A - \frac{1}{2}\log_2 e \cdot \mathbf{x}^T K^{-1}\mathbf{x}$$

- Επίσης, $\int \phi_K(\mathbf{x}) x_i x_j d\mathbf{x} = K_{ij}$. Επομένως,

$$\begin{aligned} 0 \le D(g||\phi_K) &= \int g \log\left(\frac{g}{\phi_K}\right) = -h(g) - \int g \log \phi_K \\ &\overset{(a)}{=} -h(g) - \int \phi_K \log \phi_K = -h(g) + h(\phi_K). \end{aligned}$$

(a) Προκύπτει από την παρατήρηση ότι η $\log \phi_K(\mathbf{x})$ είναι τετραγωνική μορφή (άθροισμα όρων της μορφής $a_{ij} x_i x_j$), και από την υπόθεση ότι $\int g(\mathbf{x}) x_i x_j d\mathbf{x} = \int \phi_K(\mathbf{x}) x_i x_j d\mathbf{x}$.

Σφάλμα Εκτίμησης συνεχούς τ.μ.

- Είδαμε ότι, όταν μια διακριτή τ.μ. X εκτιμάται με βάση την παρατήρηση μιας τ.μ. Y, ένα φράγμα για την πιθανότητα σφάλματος δίνεται από την ανισότητα Fano:

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

- Για εκτίμηση συνεχών τ.μ., η πιθανότητα σφάλματος δεν έχει νόημα. Πολλές φορές, για να ποσοτικοποιηθεί η επίδοση ενός εκτιμητή χρησιμοποιείται το μέσο τετραγωνικό σφάλμα (mean square error) $E(X - \hat{X})^2$.
- Θα αποδείξουμε ότι

$$E(X - \hat{X})^2 \geq \frac{1}{2\pi e} e^{2h(X)},$$

με $=$ εάν και μόνο εάν η X είναι γκαουσιανή και $\hat{X} = EX$.

Σφάλμα Εκτίμησης συνεχούς τ.μ. (συνέχεια)

$$E(X-\hat{X})^2 \geq \min_{\hat{X}} E(X-\hat{X})^2 \stackrel{(a)}{=} E(X-EX)^2$$

$$= var(X) \stackrel{(b)}{\geq} \frac{1}{2\pi e} 2^{2h(X)}.$$

(a) Ο καλύτερος εκτιμητής για τη X είναι η μέση τιμή της (b) Για δεδομένη διασπορά, η τ.μ. με τη μεγαλύτερη εντροπία είναι η Γκαουσιανή. $h(X) \leq \frac{1}{2}\log(2\pi e\sigma^2) \Rightarrow \sigma^2 \geq \frac{1}{2\pi e} 2^{2h(X)}$.

- Με χρήση του ίδιου συλλογισμού μπορούμε να δείξουμε ότι, δεδομένης τ.μ. Y,

$$E(X-\hat{X}(Y))^2 \geq \frac{1}{2\pi e} e^{2h(X|Y)}.$$

Ανακεφαλαίωση μαθήματος

- Ένας τρόπος μετάδοσης πληροφορίας πηγής μέσα από ένα κανάλι είναι πρώτα να συμπιέσουμε με βάση την κατανομή της πηγής και, στη συνέχεια, να κωδικοποίησουμε με βάση το κανάλι. Αποδεικνύεται ότι, για Διακριτά Κανάλια χωρίς μνήμη, ο τρόπος αυτός δεν υστερεί σε σχέση με οποιονδήποτε άλλον.
- Οι ποσότητες της Θεωρίας Πληροφορίας χρησιμοποιούνται (με κάποιες αλλαγές) και για την περιγραφή συνεχών τ.μ. Ομοίως, το AEP ισχύει και για συνεχείς τ.μ.
- Για δεδομένο πίνακα συσχέτισης, η κατανομή που μεγιστοποιεί τη Διαφορική Εντροπία είναι η πολυμεταβλητή Γκαουσιανή.

Προεπισκόπηση επόμενου μαθήματος

- AEP για συνεχείς τ.μ.
- Γκαουσιανό κανάλι, χωρητικότητα και Θεώρημα Κωδικοποίησης.
- Κανάλι Προσθετικού Λευκού Γκαουσιανού Θορύβου (AWGN). Χωρητικότητα.
- Παράλληλα και ανεξάρτητα κανάλια AWGN. Βέλτιστος τρόπος μετάδοσης για δεδομένη διαθέσιμη ισχύ και χωρητικότητα.
- Κανάλια Προσθετικού Έγχρωμου Γκαουσιανού Θορύβου (ACGN). Χωρητικότητα.
- Κανάλι Προσθετικού Γκαουσιανού Θορύβου με ανάδραση. Στην περίπτωση έγχρωμου θορύβου η χωρητικότητα μπορεί να αυξηθεί κατά 1/2, το πολύ, bit.