ΕΕ728
Προχωρημένα Θέματα Θεωρίας Πληροφορίας
7η διάλεξη
(2η έκδοση, 10/4/2011)

Δημήτρης-Αλέξανδρος Τουμπακάρης

Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών

5 Απριλίου 2011

Περιεχόμενα 7ης διάλεξης

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. Αντίστροφα, 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 $V^n = V_1, V_2, \dots, V_n$ 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

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

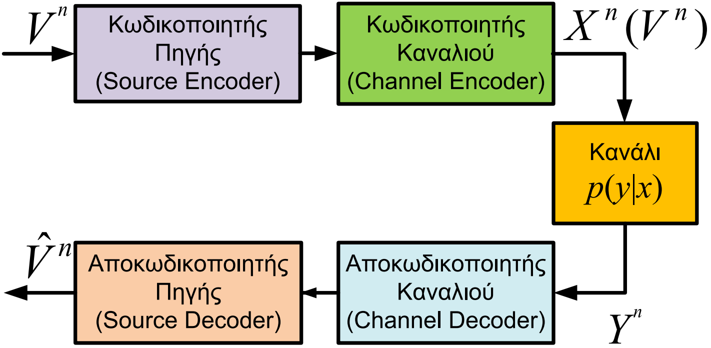
  όπου I η συνάρτηση-δείκτης και $g(\cdot)$ η συνάρτηση αποκωδικοποίησης.

- Θεώρημα 7.1. (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 $H(\mathcal{V}) < C$.

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,

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

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

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

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

Θεώρημα Διαχωρισμού Πηγής - Καναλιού
Απόδειξη αντιστρόφου (συνέχεια)

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

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

- Για $n \to \infty$, $\Pr\left\{\hat{V}^n \neq V^n\right\} \to 0$ 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. Ωστόσο, η διαφορική εντροπία είναι πιο "προβληματικό" μέγεθος σε σχέση με την εντροπία και υπάρχουν κάποιες διαφορές που θα πρέπει να προσεχτούν.
- Θα αναφερθούμε στις συνεχείς τ.μ. πιο επιγραμματικά, φροντίζοντας, όμως, να επισημαίνουμε τις διαφορές, όπου υπάρχουν.

Διαφορική Εντροπία – Ορισμός

- Ορισμός 7.2. 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 7.1. – 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.
- Η διαφορική εντροπία διακριτής τ.μ. ισούται με $-\infty$ ($2^{-\infty} = 0$).

Παράδειγμα 7.2. – Εντροπία Γκαουσιανής τ.μ.

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

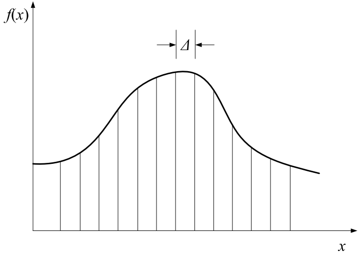
$$X \sim \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}.$$

- Με χρήση του ορισμού της διαφορικής εντροπίας,

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

Παράδειγμα 7.3. – Εντροπία διακριτής αναπαράστασης συνεχούς τ.μ.

- Έστω συνεχής τ.μ. X με συνάρτηση πυκνότητας πιθανότητας $f(x)$. Χωρίζουμε την $f(X)$ σε διαστήματα 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 7.3. – Εντροπία διακριτής αναπαράστασης συνεχούς τ.μ. (2)

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

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

Παράδειγμα 7.3. – Εντροπία διακριτής αναπαράστασης συνεχούς τ.μ. (3)

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

- Όταν $\Delta \to 0$, $H\left(X^{\Delta}\right) \to h(X) - \log \Delta$, εάν 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 (κβάντιση) της συνεχούς τ.μ. X. Επομένως, $H\left(X^{\Delta}\right) \approx h(X) + n$.
- 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 7.3.** Από κοινού διαφορική εντροπία: $h(X_1, X_2, \ldots, X_n) = -\int f(x^n) \log f(x^n) dx^n$ (εάν υπάρχει), όπου $f(x^n) = f(x_1, x_2, \ldots, x_n)$.
- **Ορισμός 7.4.** Υπό συνθήκη διαφορική εντροπία: $h(X|Y) = -\int f(x, y) \log f(x|y) dxdy$ (εάν υπάρχει).
- Όπως και στην περίπτωση διακριτών τ.μ., εάν όλες οι ποσότητες είναι πεπερασμένες,

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

Παράδειγμα 7.4. – 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 & Thomas Theorem 8.4.1) ότι

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

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

Σχετική Εντροπία και Αμοιβαία Πληροφορία για συνεχείς τ.μ.

- Ορισμός 7.5. 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.
- Ορισμός 7.6. 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

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

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.

Μεγιστοποίηση Διαφορικής Εντροπίας

- Θεώρημα 7.7. Έστω τυχαίο διάνυσμα $\mathbf{X} \in \mathbf{R}^n$ με μέση τιμή $\mathbf{m} = 0$ και πίνακα συσχέτισης $K = \mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right]$ (δηλαδή $K_{ij} = \mathbb{E}\left[X_i X_j\right],\ 1 \leq i, j \leq n$).
  Για τη διαφορική εντροπία της $\mathbf{X}$ ισχύει
  $h(\mathbf{X}) \leq \frac{1}{2}\log(2\pi e)^n|K|$, με $=$ εάν και μόνο εάν $\mathbf{X} \sim \mathcal{N}(0, K)$.
- Επομένως, για δεδομένο πίνακα συσχέτισης K, η συνεχής κατανομή που μεγιστοποιεί την εντροπία είναι η Γκαουσιανή!
- Θα μας χρησιμεύσει στον υπολογισμό της χωρητικότητας του Γκαουσιανού Καναλιού.

Μεγιστοποίηση Διαφορικής Εντροπίας (2)

- Πόρισμα 7.8. Για βαθμωτή συνεχή τ.μ. X με μέση τιμή $m = 0$ και διασπορά 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 <u>ισχύ</u> (=$\sum_i \mathbb{E}|X_i|^2 = \mathbb{E}\left[\mathbf{X}^T\mathbf{X}\right] =$ tr$\{\mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right]\} =$ tr$\{K\}$), οι πιο "αβέβαιες" είναι αυτές που ακολουθούν Γκαουσιανή κατανομή $\mathcal{N}(\mathbf{m}, K)$.

Μεγιστοποίηση Διαφορικής Εντροπίας – Απόδειξη

- Έστω $g(\mathbf{x})$ οποιαδήποτε συνάρτηση πυκνότητας πιθανότητας που ικανοποιεί τον περιορισμό συσχέτισης $\mathbb{E}_{g(\mathbf{x})}[X_i X_j] = \int g(\mathbf{x}) x_i x_j d\mathbf{x} = K_{ij}$ για όλα τα i, j. Έστω, επίσης $\phi_K(\mathbf{x})$ 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,

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

(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 $\int g(\mathbf{x}) x_i x_j d\mathbf{x} = \int \phi_K(\mathbf{x}) x_i x_j d\mathbf{x}$.

AEP για συνεχείς τ.μ.

AEP για συνεχείς τ.μ.

- Για την περίπτωση συνεχών τ.μ. το AEP μπορεί να αποδειχτεί με παρόμοιο τρόπο όπως στην περίπτωση διακριτών τυχαίων μεταβλητών. Πιο συγκεκριμένα,
    1. $\Pr\left\{A_\epsilon^{(n)}\right\} > 1 - \epsilon$ για $n > n_0$.
    2. $\mathrm{Vol}\left(A_\epsilon^{(n)}\right) \leq 2^{n(h(X)+\epsilon)}$ για όλα τα n.
    3. $\mathrm{Vol}\left(A_\epsilon^{(n)}\right) \geq (1-\epsilon)2^{n(h(X)-\epsilon)}$ για $n > n_0$.
- Παρατηρήστε ότι, για συνεχείς τ.μ., η ποσότητα που αντιστοιχεί στον αριθμό στοιχείων του τυπικού συνόλου $\left|A_\epsilon^{(n)}\right|$ είναι ο όγκος $\mathrm{Vol}(A)$ του συνόλου A:

$$\mathrm{Vol}(A) = \int_A dx_1 dx_2 \ldots dx_n.$$

AEP για συνεχείς τ.μ. (συνέχεια)

- Όσο μικρότερη είναι η εντροπία μιας συνεχούς τ.μ., τόσο μικρότερος είναι ο μέσος όγκος που καταλαμβάνει το σύνολο που περιέχει σχεδόν όλη την πιθανότητα (για δεδομένο n).
- Για n σύμβολα (διαστάσεις), $\mathrm{Vol}\left(A_\epsilon^{(n)}\right) \approx 2^{nh(X)}$. Επομένως, ο "χώρος" στον οποίο περιέχονται οι τυπικές ακολουθίες έχει πλευρά μήκους $\approx 2^{h(X)}$ *ανά διάσταση*.