Στοιχεία Θεωρίας Πληροφορίας

Αθανάσιος Π. Λιάβας

Καθηγητής

Τμήμα ΗΜΜΥ
Πολυτεχνείο Κρήτης

Εισαγωγή

Οι παρούσες σημειώσεις αποτελούν υλικό που αναπτύσσεται κατά τη διδασκαλία του μαθήματος *Στοιχεία Θεωρίας Πληροφορίας* του προγράμματος μεταπτυχιακών σπουδών του Τμήματος ΗΜΜΥ, του Πολυτεχνείου Κρήτης. Βασίζονται σε πολύ μεγάλο βαθμό στα βιβλία

1. T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley, 2006.
2. R. W. Yeung. *Information Theory and Network Coding.* Springer, 2008.
3. R. McEliece. *The Theory of Information and Coding.* Cambridge, 2004.

Επίσης χρήσιμα φάνηκαν τα βιβλία

3. D. Tse and P. Viswanath. *Fundamentals of Wireless Communications.* Cambridge University Press, 2005.
4. R. Gallager. *Information Theory and Reliable Communication.* Wiley, 1968.
5. R. Ash. *Information Theory.* Dover, 1990.
6. S. Ihara. *Information Theory for Continuous Systems.* World Scientific, 1993.

Θα ήθελα να ευχαριστήσω τους φοιτητές που συνέβαλαν στη βελτίωση των σημειώσεων με ερωτήσεις, απορίες και παρατηρήσεις.

Οι σημειώσεις είναι αφιερωμένες στον ιδρυτή της περιοχής της Θεωρίας Πληροφορίας, τον Claude Elwood Shannon, ο οποίος, κατά τη γνώμη μου, υπήρξε ένας από τους πιο σημαντικούς επιστήμονες του εικοστού αιώνα.

Τμήμα ΗΜΜΥ, Χανιά, Εαρινό Εξάμηνο 2012–2013

Αθανάσιος Π. Λιάβας

Επισημαίνω ότι οι παρούσες σημειώσεις δεν έχουν υποστεί εξαντλητικό έλεγχο και, συνεπώς, ίσως περιέχουν τυπογραφικά (και όχι μόνο) λάθη. Τα σχόλια και οι παρατηρήσεις σας είναι ευπρόσδεκτα.

Κεφάλαιο 1

Εντροπία, Αμοιβαία Πληροφορία

Η έννοια της πληροφορίας είναι ευρεία και δεν μπορεί να αποδοθεί πλήρως από ένα ορισμό.

Σε αυτό το κεφάλαιο, για κάθε συνάρτηση μάζας πιθανότητας ορίζουμε την εντροπία, η οποία έχει πολλές ιδιότητες που συμφωνούν με τη διαισθητική έννοια του μέτρου πληροφορίας. Η έννοια της εντροπίας επεκτείνεται στην αμοιβαία πληροφορία, η οποία είναι ένα μέτρο της πληροφορίας που παρέχει μία τυχαία μεταβλητή για μια άλλη.

Όπως ορίζονται σε αυτό το κεφάλαιο, η εντροπία και η αμοιβαία πληροφορία, θα μπορούσαν να είναι, απλώς, δύο ενδιαφέροντες ορισμοί ποσοτήτων, οι οποίες είναι δύσκολο να οριστούν με πλήρη σαφήνεια. Για παράδειγμα, ποιος είναι ο σαφής ορισμός της έννοιας "πληροφορία;" Όπως θα δούμε σε επόμενα κεφάλαια, η μεγάλη λειτουργική σημασία της εντροπίας και της αμοιβαίας πληροφορίας έγγειται στο ότι καθορίζουν με ακρίβεια τα όρια της "συμπίεσης δεδομένων" και του "ρυθμού αξιόπιστης επικοινωνίας."

1.1 Εντροπία

Έστω X διακριτή τυχαία μεταβλητή (δτμ), η οποία παίρνει τιμές από το πεπερασμένο ή αριθμήσιμο σύνολο $\mathcal{X}$, με συνάρτηση μάζας πιθανότητας (σμπ) $p_X(x) = \Pr(X = x)$, για $x \in \mathcal{X}$ (συνήθως, όταν η δτμ υπονοείται, για απλούστευση συμβολισμού, θα γράφουμε $p(x)$ αντί για $p_X(x)$).

Ορισμός: Η **εντροπία** (entropy) $H(X)$ της δτμ X, η οποία λαμβάνει τιμές από το σύνολο

$\mathcal{X}$ με συνάρτηση μάζας πιθανότητας $p_X(x)$, ορίζεται ως εξής:

$$H(X) := -\sum_{x\in\mathcal{X}} p_X(x)\log p_X(x). \tag{1.1}$$

Σχόλιο 1: Στη σχέση (1.1), παραθέτουμε ένα μαθηματικό ορισμό χωρίς να δίνουμε καμία εξήγηση για το πώς φθάσαμε σε αυτόν. Ένας τρόπος να καταλήξουμε σε αυτό τον ορισμό είναι η λεγόμενη αξιωματική προσέγγιση, κατά την οποία θέτουμε ορισμένους περιορισμούς τους οποίους, λογικά, θα πρέπει να ικανοποιεί κάθε μέγεθος που εκφράζει "ποσό πληροφορίας". Για παράδειγμα, είναι λογικό να θέσουμε ως περιορισμό το ότι "η συνολική πληροφορία που είναι προσεταιρισμένη με ανεξάρτητα γεγονότα είναι το άθροισμα των επιμέρους πληροφοριών." Αυτή την προσέγγιση ακολούθησε ο Shannon στην αρχική του εργασία το 1948.[1] Σε αυτές τις σημειώσεις, δεν θα ασχοληθούμε με την αξιωματική προσέγγιση. □

Σχόλιο 2: Μία διαισθητικά ικανοποιητική προσέγγιση του ορισμού της εντροπίας έχει ως εξής. Έστω η τυχαία μεταβλητή

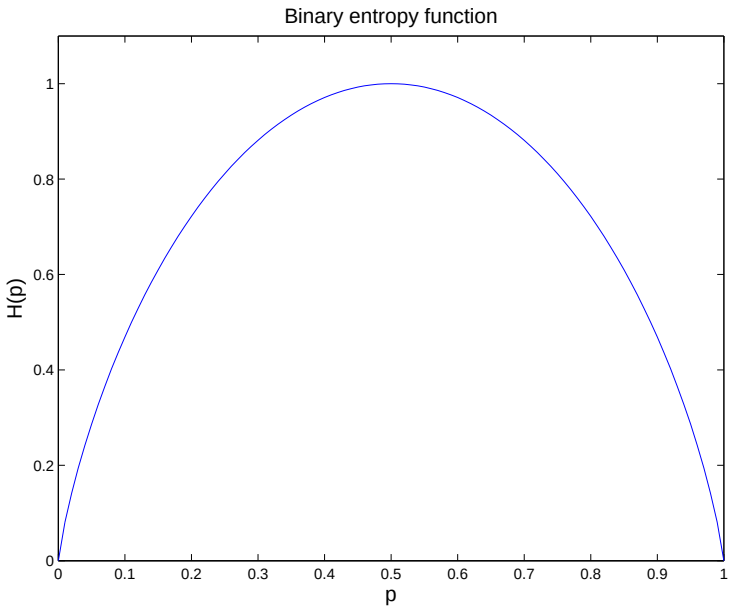
$$I(X) = -\log p(X).$$

Για $x \in \mathcal{X}$, η ποσότητα $I(x)$ μπορεί να θεωρηθεί ως ένα "μέτρο της πληροφορίας" που είναι προσεταιρισμένη με το ενδεχόμενο x. Αν η πιθανότητα του ενδεχομένου x είναι "μικρή," δηλαδή, $p(x) \approx 0$, τότε $I(x) \gg 1$, δηλαδή, η πληροφορία που είναι προσεταιρισμένη με την εμφάνισή του x κατά την υλοποίηση του πειράματος είναι "μεγάλη." Αν η πιθανότητα του ενδεχομένου x είναι "μεγάλη," δηλαδή, $p(x) \approx 1$, τότε $I(x) \approx 0$, δηλαδή, η πληροφορία που είναι προσεταιρισμένη με την εμφάνισή του x κατά την υλοποίηση του πειράματος είναι "μικρή." Η εντροπία $H(X)$ μπορεί να θεωρηθεί ως η μέση πληροφορία που λαμβάνουμε με την παρατήρηση της τυχαίας μεταβλητής X. □

Αν η βάση του λογαρίθμου είναι το 2, τότε η εντροπία εκφράζεται σε bits, το οποίο αποτελεί συντομογραφία του binary digits. Αν η βάση του λογαρίθμου είναι το e, τότε η εντροπία εκφράζεται σε nats, το οποίο αποτελεί συντομογραφία του natural units.

Παρατήρηση 1: Ορίζουμε $0\log 0 = 0$, διότι $x\log x \to 0$ όταν $x \to 0$.

[1] Η εργασία αυτή αξίζει να διαβαστεί ακόμα και σήμερα.



Σχήμα 1.1: Εντροπία $H(p)$ δυαδικής τυχαίας μεταβλητής.

Παρατήρηση 2: Η εντροπία είναι **συναρτησιακό** (functional) της συνάρτησης μάζας πιθανότητας, $p(x)$, της δτμ X. Η τιμή της εντροπίας δεν εξαρτάται από τις τιμές της X αλλά από τις πιθανότητες των τιμών! Για αυτό το λόγο, αρκετές φορές η εντροπία συμβολίζεται ως $H(p_X)$ αντί για $H(X)$.

Παρατήρηση 3: Η μέση τιμή της συνάρτησης $g(X)$ της δτμ X ορίζεται ως

$$\mathcal{E}_{p_X}[g(X)] := \sum_{x\in\mathcal{X}} p_X(x)g(x).$$

Συνεπώς, η εντροπία μπορεί να συμβολιστεί ως

$$H(X) = \mathcal{E}p_X\left[\log\frac{1}{p(X)}\right] = -\mathcal{E}_{p_X}\left[\log p(X)\right].$$

Παράδειγμα: Έστω η τυχαία μεταβλητή X με τιμές στο σύνολο $\{0,1\}$ και συνάρτηση μάζας πιθανότητας $p_X(0) = p$ και $p_X(1) = 1-p$. Η εντροπία της X είναι

$$H(X) = -p\log p - (1-p)\log(1-p). \tag{1.2}$$

Η $H(X)$ σχεδιάζεται στο Σχήμα 1.1. Για $p = 0$ ή $p = 1$, η $H(X)$ παίρνει την τιμή μηδέν, διότι η X παίρνει πάντα την τιμή 1 ή 0, αντίστοιχα. Η μέγιστη τιμή της $H(X)$ επιτυγχάνεται

για $p = \frac{1}{2}$ και είναι ίση με 1. Συνεπώς, μία δυαδική τυχαία μεταβλητή "μεταφέρει" πληροφορία το πολύ 1 bit, με το μέγιστο να επιτυγχάνεται όταν οι τιμές 0 και 1 είναι ισοπίθανες. □

Λήμμα: *Η εντροπία είναι μη αρνητική.*

$$H(X) \geq 0. \tag{1.3}$$

Απόδειξη: Η σχέση $0 \leq p(x) \leq 1$ συνεπάγεται ότι $\log \frac{1}{p(x)} = -\log p(x) \geq 0$, το οποίο οδηγεί εύκολα στην απόδειξη του λήμματος. □

Λήμμα: $H_b(X) = H_a(X) \log_b a$.
Απόδειξη: Η απόδειξη βασίζεται στη σχέση $\log_b p(x) = \log_a p(x) \log_b a$. □

1.2 Από-κοινού και υπό-συνθήκη εντροπία

Ορισμός: Η **από-κοινού εντροπία** (joint entropy) $H(X,Y)$ του ζεύγους δτμ (X,Y), με από-κοινού συνάρτηση μάζας πιθανότητας $p(x,y)$, για $x \in \mathcal{X}$ και $y \in \mathcal{Y}$, ορίζεται ως εξής:

$$H(X,Y) := -\mathcal{E}_{p_{X,Y}} [\log p(X,Y)] = -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log p(x,y). \tag{1.4}$$

Η **υπό-συνθήκη εντροπία** (conditional entropy) $H(Y|X)$ ορίζεται ως εξής. Αρχικά, υποθέτουμε ότι $X = x$. Συνεπώς, η κατανομή της Y είναι $p_{Y|X}(y|x)$ και η εντροπία της Y είναι

$$H(Y|X = x) := -\sum_{y\in\mathcal{Y}} p(y|x) \log p(y|x).$$

Η υπό-συνθήκη εντροπία $H(Y|X)$ είναι η μέση τιμή των $H(Y|X = x)$, δηλαδή

$$\begin{aligned} H(Y|X) &:= \sum_{x\in\mathcal{X}} p(x) H(Y|X = x) \\ &= -\sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} p(y|x) \log p(y|x) \\ &= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log p(y|x) \\ &= -\mathcal{E}_{p_{X,Y}} [\log p(Y|X)] . \end{aligned} \tag{1.5}$$

Θεώρημα: (Κανόνας Αλυσίδας)

$$\begin{aligned} H(X,Y) &= H(X) + H(Y|X) \\ &= H(Y) + H(X|Y). \end{aligned} \tag{1.6}$$

Απόδειξη:

$$\begin{aligned} H(X,Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y) \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left[p(x)p(y|x)\right] \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y|x) \\ &= -\sum_{x\in\mathcal{X}} p(x)\log p(x) + H(Y|X) = H(X) + H(Y|X). \end{aligned}$$

Με τον ίδιο τρόπο αποδεικνύεται η δεύτερη ισότητα. □

Διαισθητικά, η παραπάνω σχέση δηλώνει ότι η αβεβαιότητα του ζεύγους τυχαίων μεταβλητών (X,Y) μπορεί να αρθεί σταδιακά ή, ισοδύναμα, η πληροφορία που παρέχει το ζεύγος (X,Y) μπορεί να αποκαλυφθεί σταδιακά. Αρχικά, αποκαλύπτεται η X και κατόπιν, υπό τη γνώση της X, αποκαλύπτεται η Y, και αντίστροφα.

Πόρισμα: $H(X,Y|Z) = H(X|Z) + H(Y|X,Z)$.
Απόδειξη: Η απόδειξη αφήνεται σαν άσκηση. □

1.3 Σχετική Εντροπία και Αμοιβαία Πληροφορία

Ορισμός: Η **σχετική εντροπία** (relative entropy) ή Kullback-Leibler απόσταση μεταξύ δύο συναρτήσεων μάζας πιθανότητας $p(x)$ και $q(x)$ ορίζεται ως εξής:

$$D(p\|q) := \mathcal{E}_p \log\frac{p(X)}{q(X)} = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}. \tag{1.7}$$

Όπως θα αποδείξουμε στη συνέχεια, η σχετική εντροπία είναι μη-αρνητική και ισούται με μηδέν αν, και μόνο αν, $p=q$. Χρησιμοποιείται συνήθως για να "μετρήσει την απόσταση" μεταξύ των συναρτήσεων μάζας πιθανότητας $p(x)$ και $q(x)$, παρόλο που δεν αποτελεί **μετρική**

(metric) διότι δεν είναι συμμετρική και δεν ικανοποιεί την τριγωνική ανισότητα.

Ορισμός: Έστω διακριτές τ.μ. X και Y με από κοινού συνάρτηση μάζας πιθανότητας $p(x,y)$ και περιθώριες κατανομές $p(x)$ και $p(y)$. Ως **αμοιβαία πληροφορία** (mutual information) $I(X;Y)$ ορίζουμε τη σχετική εντροπία ανάμεσα στην από-κοινού κατανομή και στο γινόμενο των περιθώριων κατανομών, δηλαδή

$$\begin{aligned} I(X;Y) &:= D(p(x,y)\|p(x)p(y)) \\ &= \mathcal{E}\left[\log \frac{p(X,Y)}{p(X)p(Y)}\right] \\ &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}. \end{aligned} \tag{1.8}$$

1.4 Σχέση Αμοιβαίας Πληροφορίας και Εντροπίας

Η αμοιβαία πληροφορία και η εντροπία συνδέονται ως εξής:

$$\begin{aligned} I(X,Y) &= \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\ &= \sum_{x,y} p(x,y)\log\frac{p(x|y)}{p(x)} \\ &= -\sum_{x,y} p(x,y)\log p(x) + \sum_{x,y} p(x,y)\log p(x|y) \\ &= -\sum_{x} p(x)\log p(x) - \left(-\sum_{x,y} p(x,y)\log p(x|y)\right) \\ &= H(X) - H(X|Y). \end{aligned} \tag{1.9}$$

Συνεπώς, η αμοιβαία πληροφορία $I(X;Y)$ ισούται με τη μείωση της αβεβαιότητας της X εξαιτίας της γνώσης της Y. Εύκολα αποδεικνύεται ότι $I(X;Y) = I(Y;X)$.

Θεώρημα: Οι παρακάτω σχέσεις ισχύουν:

$$\begin{aligned} &I(X;Y) = H(X) + H(Y) - H(X,Y) = H(Y) - H(Y|X) = I(Y;X) \\ &I(X;X) = H(X). \end{aligned} \tag{1.10}$$

Απόδειξη: Η απόδειξη αφήνεται σαν άσκηση. □

1.5 Κανόνες Αλυσίδας

Θεώρημα: Έστω $X_1, \ldots, X_n$ τ.μ. με συνάρτηση μάζας πιθανότητας $p(x_1, \ldots, x_n)$. Τότε $H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1)$.
Απόδειξη: Θα χρησιμοποιήσουμε τη σχέση

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i|x_{i-1}, \ldots, x_1).$$

$$\begin{aligned}
H(X_1, \ldots, X_n) &= -\sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n) \\
&= -\sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \log \prod_{i=1}^{n} p(x_i|x_{i-1}, \ldots, x_1) \\
&= -\sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \sum_{i=1}^{n} \log p(x_i|x_{i-1}, \ldots, x_1) \\
&= -\sum_{i=1}^{n} \sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \log p(x_i|x_{i-1}, \ldots, x_1) \\
&= -\sum_{i=1}^{n} \sum_{x_1,\ldots,x_i} p(x_1, \ldots, x_i) \log p(x_i|x_{i-1}, \ldots, x_1) \\
&= \sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1).
\end{aligned}$$

□

Ορισμός: Η υπό-συνθήκη αμοιβαία πληροφορία των τ.μ. X και Y δεδομένης της τ.μ. Z ορίζεται ως εξής:

$$I(X;Y|Z) := \mathcal{E}_{p_{X,Y,Z}} \log \frac{p(X,Y|Z)}{p(X|Z)p(Y|Z)} = H(X|Z) - H(X|Y,Z). \tag{1.11}$$

Θεώρημα: Κανόνας αλυσίδας για την αμοιβαία πληροφορία.

$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y|X_{i-1}, \ldots, X_1). \tag{1.12}$$

Απόδειξη:

$$\begin{aligned} I(X_1,\ldots,X_n;Y) &= H(X_1,\ldots,X_n) - H(X_1,\ldots,X_n|Y) \\ &= \sum_{i=1}^{n} H(X_i|X_{i-1},\ldots,X_1) - \sum_{i=1}^{n} H(X_i|X_{i-1},\ldots,X_1,Y) \\ &= \sum_{i=1}^{n} I(X_i;Y|X_{i-1},\ldots,X_1). \end{aligned} \tag{1.13}$$

□

1.6 Ανισότητα Jensen

Ορισμός: Η συνάρτηση $f(x)$ καλείται **κυρτή** (convex) στο διάστημα (a,b) αν για κάθε $x_1, x_2 \in (a,b)$ και $0 \leq \lambda \leq 1$ ισχύει

$$f\left(\lambda x_1 + (1-\lambda)x_2\right) \leq \lambda f(x_1) + (1-\lambda) f(x_2). \tag{1.14}$$

Η συνάρτηση f καλείται αυστηρά κυρτή αν η ισότητα ισχύει μόνο για $\lambda = 0, 1$. Η συνάρτηση f καλείται **κοίλη** (concave) αν η $-f$ είναι κυρτή.

Παρατήρηση: Η γραμμική συνάρτηση $f(x) = ax + b$ είναι ταυτόχρονα κυρτή και κοίλη.

Θεώρημα: Αν η συνάρτηση f έχει δεύτερη παράγωγο η οποία είναι μη-αρνητική (θετική) παντού, τότε είναι κυρτή (αυστηρά κυρτή).
Απόδειξη: Βλέπε βιβλία Απειροστικού Λογισμού. □

Συμβολίζουμε με E τη μέση τιμή $EX = \sum_{x\in\mathcal{X}} xp(x)$ για διακριτή τ.μ. και $EX = \int xf(x)dx$ για συνεχή τυχαία μεταβλητή.

Θεώρημα: *Ανισότητα Jensen.* Αν η f είναι κυρτή συνάρτηση και X τ.μ., τότε

$$Ef(X) \geq f(EX). \tag{1.15}$$

Επιπλέον, αν η f είναι αυστηρά κυρτή, τότε ισότητα στην (1.15) συνεπάγεται $X = EX$ με πιθανότητα 1, δηλαδή, X σταθερά, με πιθανότητα 1.

Απόδειξη: Θα δώσουμε την απόδειξη μόνο για διακριτές τ.μ., η οποία γίνεται με επαγωγή ως προς το πλήθος των σημείων μάζας πιθανότητας. Για την απόδειξη για συνεχείς τυχαίες μεταβλητές, δείτε κάποιο καλό βιβλίο Πραγματικής Ανάλυσης.

Για κατανομή με δύο σημεία μάζας πιθανότητας, p_1 και $p_2 = 1 - p_1$, η ανισότητα είναι

$$p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 x_1 + p_2 x_2), \tag{1.16}$$

η οποία ισχύει από τον ορισμό της κυρτής συνάρτησης. Υποθέτουμε ότι το θεώρημα αληθεύει για κατανομές με $(k-1)$ σημεία μάζας πιθανότητας. Δηλαδή, για οποιαδήποτε κατανομή $(q_1, \ldots, q_{k-1})$ έχουμε

$$\sum_{i=1}^{k-1} q_i f(x_i) \geq f\left(\sum_{i=1}^{k-1} q_i x_i\right). \tag{1.17}$$

Έστω κατανομή k σημείων $(p_1, \ldots, p_k)$ και $p'_i = p_i/(1-p_k)$, για $i = 1, \ldots, k-1$. Παρατηρήστε ότι $(p'_1, \ldots, p'_{k-1})$ είναι κατανομή $(k-1)$ σημείων. Τότε

$$\begin{aligned}
\sum_{i=1}^{k} p_i f(x_i) &\stackrel{!}{=} p_k f(x_k) + (1-p_k) \sum_{i=1}^{k-1} p'_i f(x_i) \\
&\stackrel{(1.17)}{\geq} p_k f(x_k) + (1-p_k) f\left(\sum_{i=1}^{k-1} p'_i x_i\right) \\
&\stackrel{(1.16)}{\geq} f\left(p_k x_k + (1-p_k) \sum_{i=1}^{k-1} p'_i x_i\right) \\
&= f\left(\sum_{i=1}^{k} p_i x_i\right).
\end{aligned} \tag{1.18}$$

□

Θεώρημα: *Ανισότητα Πληροφορίας* (Information inequality). Έστω $p(x)$ και $q(x)$, με $x \in \mathcal{X}$, συναρτήσεις μάζας πιθανότητας. Τότε

$$D(p\|q) \geq 0 \tag{1.19}$$

με ισότητα αν και μόνο αν $p(x) = q(x)$, $\forall x \in \mathcal{X}$.
Απόδειξη α: Έστω $A = \{x : p(x) > 0\}$. Τότε

$$\begin{aligned} -D(p\|q) &= -\sum_{x\in A} p(x) \log \frac{p(x)}{q(x)} \\ &= \sum_{x\in A} p(x) \log \frac{q(x)}{p(x)} \quad (\log t : \text{ αυστηρά κοίλη συνάρτηση του } t) \\ &\overset{\text{J.I.}}{\leq} \log \sum_{x\in A} p(x) \frac{q(x)}{p(x)} \\ &= \log \sum_{x\in A} q(x) \\ &\leq \log \sum_{x\in \mathcal{X}} q(x) = \log 1 = 0. \end{aligned} \tag{1.20}$$

Έχουμε ισότητα στην (1.19) αν και μόνο αν $q(x)/p(x) = 1$, $\forall x \in \mathcal{X}$. Άρα, $D(p\|q) = 0$ αν, και μόνο αν, $q(x) = p(x)$, $\forall x \in \mathcal{X}$.

Απόδειξη β: Ξεκινάμε από τις ταυτότητες:

$$\sum_{x\in\mathcal{X}} p(x) = \sum_{x\in\mathcal{X}} q(x) = 1.$$

Θα αποδείξουμε ότι

$$-\sum_{x\in\mathcal{X}} p(x) \log p(x) \leq -\sum_{x\in\mathcal{X}} p(x) \log q(x),$$

με ισότητα αν, και μόνο αν, $p(x) = q(x)$, $\forall x \in \mathcal{X}$. Αφού $\log_2 x = \log_2 e \ln x$, το αποτέλεσμα είναι αμετάβλητο από την αλλαγή βάσης λογαρίθμου. Χρησιμοποιούμε τη βασική ανισότητα

$$\ln x \leq x - 1$$

στην οποία η ισότητα ισχύει αν, και μόνο αν, $x = 1$. Τότε

$$\ln\left(\frac{q(x)}{p(x)}\right) \leq \frac{q(x)}{p(x)} - 1$$

με ισότητα αν και μόνο αν $p(x) = q(x)$. Πολλαπλασιάζοντας με $p(x)$ και αθροίζοντας έχουμε:

$$\sum_{x\in\mathcal{X}} p(x) \ln\left(\frac{q(x)}{p(x)}\right) \leq \sum_{x\in\mathcal{X}} (q(x) - p(x)) = 1 - 1 = 0,$$

με ισότητα αν και μόνο αν $p(x) = q(x)$, $\forall x \in \mathcal{X}$. □

Πόρισμα: Για οποιεσδήποτε τ.μ. X και Y ισχύει

$$I(X;Y) \geq 0 \tag{1.21}$$

με ισότητα αν, και μόνο αν, οι X και Y είναι ανεξάρτητες.
Απόδειξη: $I(X;Y) = D(p(x,y)\|p(x)\,p(y)) \geq 0$, με ισότητα αν και μόνο αν $p(x,y) = p(x)\,p(y)$. □

Πόρισμα: $I(X;Y|Z) \geq 0$.

Θεώρημα: $H(X) \leq \log|\mathcal{X}|$, όπου $|\mathcal{X}|$ συμβολίζει τον πληθάριθμο του συνόλου $\mathcal{X}$, με ισότητα αν, και μόνο αν, η X έχει ομοιόμορφη κατανομή στο $\mathcal{X}$.

Απόδειξη: Έστω $u(x) = \frac{1}{|\mathcal{X}|}$ η συνάρτηση μάζας πιθανότητας της ομοιόμορφης κατανομής στο $\mathcal{X}$ και $p(x)$ η συνάρτηση μάζας πιθανότητας της X. Τότε

$$0 \leq D(p\|u) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{u(x)} = \sum_{x\in\mathcal{X}} p(x)\log|\mathcal{X}| + \sum_{x\in\mathcal{X}} p(x)\log p(x) = \log|\mathcal{X}| - H(X).$$

□

Θεώρημα: *Η συνθήκη δεν αυξάνει την εντροπία* (Conditioning does not increase entropy).

$$H(X|Y) \leq H(X) \tag{1.22}$$

με ισότητα αν και μόνο αν X και Y ανεξάρτητες.

Απόδειξη: $0 \leq I(X;Y) = H(X) - H(X|Y)$. □

Θεώρημα: Έστω $X_1, \ldots, X_n$ τ.μ. με σμπ $p(x_1, \ldots, x_n)$. Τότε

$$H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i) \tag{1.23}$$

με ισότητα αν, και μόνο αν, τα X_i είναι ανεξάρτητα μεταξύ τους.

Απόδειξη:

$$H(X_1,\ldots,X_n)=\sum_{i=1}^{n}H(X_i|X_{i-1},\ldots,X_1)\leq\sum_{i=1}^{n}H(X_i).$$

□

1.7 Ανισότητα log sum

Η ανισότητα που θα μελετήσουμε είναι απλή συνέπεια της κοιλότητας του λογαρίθμου.

Θεώρημα: *Ανισότητα log sum.* Για μη αρνητικούς αριθμούς $a_1,\ldots,a_n$ και $b_1,\ldots,b_n$ ισχύει

$$\sum_{i=1}^{n}a_i\log\frac{a_i}{b_i}\geq\left(\sum_{i=1}^{n}a_i\right)\log\frac{\sum_{i=1}^{n}a_i}{\sum_{i=1}^{n}b_i} \tag{1.24}$$

(Εξαιτίας της συνέχειας της $x\log x$, ορίζουμε $0\log 0=0, a\log\frac{0}{0}=0$).

Απόδειξη: Χωρίς βλάβη της γενικότητας, υποθέτουμε ότι $a_i>0$ και $b_i>0$. Η συνάρτηση $f(t)=t\log t$ είναι αυστηρά κυρτή για $t>0$, διότι $f''(t)=\frac{1}{t}\log e>0$ για $t>0$. Από την ανισότητα Jensen, έχουμε

$$\sum_i a_i f(t_i)\geq f\left(\sum_i a_i t_i\right) \tag{1.25}$$

για $a_i\geq 0,\sum_i a_i=1$. Θέτοντας $a_i=\frac{b_i}{\sum_j b_j}$ και $t_i=a_i/b_i$, η (1.25) γίνεται (μετά από απλές πράξεις)

$$\sum_i\frac{a_i}{\sum_j b_j}\log\left(\frac{a_i}{b_i}\right)\geq\sum_i\frac{a_i}{\sum_j b_j}\log\left(\sum_i\frac{a_i}{\sum_j b_j}\right). \tag{1.26}$$

Χρησιμοποιώντας τη σχέση

$$\sum_i\frac{a_i}{\sum_j b_j}=\frac{\sum_i a_i}{\sum_i b_i} \tag{1.27}$$

λαμβάνουμε

$$\sum_i\frac{a_i}{\sum_j b_j}\log\frac{a_i}{b_i}\geq\sum_i\frac{a_i}{\sum_j b_j}\log\frac{\sum_i a_i}{\sum_j b_j}. \tag{1.28}$$

Εξάγοντας τον όρο $\sum_j b_j(>0)$, στον παρονομαστή των πρώτων κλασμάτων των αθροισμάτων, κοινό παράγοντα, αποδεικνύουμε την ανισότητα (1.24). □

Θεώρημα: Η $D(p\|q)$ είναι κυρτή συνάρτηση του ζεύγους (p, q). Δηλαδή, αν (p_1, q_1) και (p_2, q_2) δύο ζεύγη συναρτήσεων μάζας πιθανότητας, τότε, για $0 \leq \lambda \leq 1$, ισχύει:

$$D(\lambda p_1 + (1-\lambda)p_2 \| \lambda q_1 + (1-\lambda)q_2) \leq \lambda D(p_1\|q_1) + (1-\lambda)D(p_2\|q_2). \tag{1.29}$$

Απόδειξη: Από την ανισότητα (1.24) λαμβάνουμε:

$$\begin{aligned}[\lambda p_1(x) + (1-\lambda)p_2(x)] \log \frac{\lambda p_1(x) + (1-\lambda)p_2(x)}{\lambda q_1(x) + (1-\lambda)q_2(x)} & \\ \leq \lambda p_1(x) \log \frac{\lambda p_1(x)}{\lambda q_1(x)} + (1-\lambda)p_2(x) \log \frac{(1-\lambda)p_2(x)}{(1-\lambda)q_2(x)} & \\ = \lambda p_1(x) \log \frac{p_1(x)}{q_1(x)} + (1-\lambda)p_2(x) \log \frac{p_2(x)}{q_2(x)}. &\end{aligned} \tag{1.30}$$

Αθροίζοντας για όλα τα x λαμβάνουμε την (1.29). □

Θεώρημα: Η εντροπία $H(p)$ είναι κοίλη συνάρτηση της συνάρτησης μάζας πιθανότητας $p(x)$.

Απόδειξη: Από τον ορισμό της σχετικής εντροπίας, έχουμε ότι $H(p) = \log|\mathcal{X}| - D(p\|u)$, όπου u είναι η ομοιόμορφη κατανομή με $|\mathcal{X}|$ ενδεχόμενα. Η κοιλότητα της H ως προς την $p(x)$ έπεται από την κυρτότητα της D ως προς την $p(x)$. □

Θεώρημα: Έστω $(X, Y) \sim p(x, y) = p(x)\, p(y|x)$. Για δεδομένη $p(y|x)$, η αμοιβαία πληροφορία $I(X; Y)$ είναι κοίλη συνάρτηση της $p(x)$.

Απόδειξη: Υποθέτουμε ότι η συνάρτηση μεταφοράς του καναλιού $p(y|x)$ είναι δεδομένη. Έστω $0 \leq \lambda \leq 1$ και είσοδοι X_1, X_2 και X με σμπ, αντίστοιχα, $p_1(x)$, $p_2(x)$ και $p(x) = \lambda p_1(x) + (1-\lambda)p_2(x)$. Οι έξοδοι συμβολίζονται με Y_1, Y_2 και Y, αντίστοιχα. Οι αντίστοιχες από-κοινού σμπ είναι $p_1(x, y) = p_1(x)p(y|x)$, $p_2(x, y) = p_2(x)p(y|x)$ και

$$p(x, y) = p(x)p(y|x) = (\lambda p_1(x) + (1-\lambda)p_2(x))p(y|x) = \lambda p_1(x, y) + (1-\lambda)p_2(x, y).$$

Τότε,

$$\begin{aligned}
&\lambda I(X_1;Y_1) + (1-\lambda) I(X_2;Y_2) - I(X;Y) \\
&\quad = \lambda \sum_{x,y} p_1(x,y) \log \frac{p(y|x)}{p_1(y)} + (1-\lambda) \sum_{x,y} p_2(x,y) \log \frac{p(y|x)}{p_2(y)} - \sum_{x,y} p(x,y) \log \frac{p(y|x)}{p(y)} \\
&\quad = \lambda \sum_{x,y} p_1(x,y) \log \frac{p(y)}{p_1(y)} + (1-\lambda) \sum_{x,y} p_2(x,y) \log \frac{p(y)}{p_2(y)}.
\end{aligned}$$

Στη συνέχεια, εφαρμόζουμε την ανισότητα Jensen σε καθένα από τα δύο αθροίσματα. Τότε, για $i = 1, 2$,

$$\sum_{x,y} p_i(x,y) \log \frac{p(y)}{p_i(y)} \leq \log \left(\sum_{x,y} p_i(x,y) \frac{p(y)}{p_i(y)} \right) = \log \left(\sum_{y} p(y) \right) = 0. \tag{1.31}$$

Συνεπώς,

$$I(X;Y) \geq \lambda I(X_1;Y_1) + (1-\lambda) I(X_2;Y_2) \tag{1.32}$$

αποδεικνύοντας ότι, για δεδομένη συνάρτηση μεταφοράς καναλιού $p(y|x)$, η αμοιβαία πληροφορία είναι κοίλη συνάρτηση της σμπ της εισόδου $p(x)$. □

Θεώρημα: Έστω $(X,Y) \sim p(x,y) = p(x)\, p(y|x)$. Για δεδομένη $p(x)$, η αμοιβαία πληροφορία $I(X;Y)$ είναι κυρτή συνάρτηση της $p(y|x)$.

Απόδειξη: Υποθέτουμε ότι η σμπ της εισόδου είναι $p(x)$. Έστω κανάλια με συναρτήσεις μεταφοράς $p_1(y|x)$, $p_2(y|x)$ και $p(y|x) = \lambda p_1(y|x) + (1-\lambda) p_2(y|x)$, με $0 \leq \lambda \leq 1$. Θα πρέπει να αποδείξουμε ότι

$$I(X;Y) \leq \lambda I(X;Y_1) + (1-\lambda) I(X;Y_2). \tag{1.33}$$

Υπολογίζουμε την ποσότητα

$$\begin{aligned}
&I(X;Y) - \lambda I(X;Y_1) - (1-\lambda) I(X;Y_2) \\
&\quad = \sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)} - \lambda \sum_{x,y} p_1(x,y) \log \frac{p_1(x|y)}{p(x)} - (1-\lambda) \sum_{x,y} p_2(x,y) \log \frac{p_2(x|y)}{p(x)} \\
&\quad = \lambda \sum_{x,y} p_1(x,y) \log \frac{p(x|y)}{p_1(x|y)} + (1-\lambda) \sum_{x,y} p_2(x,y) \log \frac{p(x|y)}{p_2(x|y)}.
\end{aligned}$$

Εφαρμόζοντας την ανισότητα Jensen, για $i = 1, 2$, λαμβάνουμε

$$\begin{aligned}
\sum_{x,y} p_i(x,y) \log \frac{p(x|y)}{p_i(x|y)} &\leq \log \left(\sum_{x,y} p_i(x,y) \frac{p(x|y)}{p_i(x|y)} \right) \\
&= \log \left(\sum_{x,y} p_i(y) p_i(x|y) \frac{p(x|y)}{p_i(x|y)} \right) \\
&= \log \left(\sum_{x,y} p_i(y) p(x|y) \right) \\
&= \log \left(\sum_{y} p_i(y) \sum_{x} p(x|y) \right) \\
&= 0.
\end{aligned} \tag{1.34}$$

Η σχέση (1.33) αποδείχθηκε. □

1.8 Ανισότητα επεξεργασίας δεδομένων

Ορισμός: Οι τυχαίες μεταβλητές X, Y, Z σχηματίζουν αλυσίδα Markov αν η υπό συνθήκη κατανομή της Z εξαρτάται μόνο από την Y και όχι από την X, δηλ. $p(Z|Y, X) = p(Z|Y)$. Το γεγονός αυτό συμβολίζεται ως $X \to Y \to Z$. Πιο συγκεκριμένα, $X \to Y \to Z$ αν

$$p(x, y, z) = p(x)\, p(y|x)\, p(z|y). \tag{1.35}$$

Απλές συνέπειες του ορισμού είναι τα εξής:

1. $X \to Y \to Z$ αν, και μόνο αν, X και Z ανεξάρτητες δεδομένης της Y, δηλαδή

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)} = \frac{p(x, y)\, p(z|y, x)}{p(y)} = \frac{p(x, y)\, p(z|y)}{p(y)} = p(x|y)\, p(z|y).$$

2. $X \to Y \to Z$ συνεπάγεται $Z \to Y \to X$, και η συνθήκη αλυσίδας Markov εκφράζεται ως $X \leftrightarrow Y \leftrightarrow Z$.

3. Αν η Z είναι συνάρτηση της Y, έστω $Z = f(Y)$, τότε η γνώση της Y καθορίζει πλήρως την Z. Συνεπώς, για κάθε τυχαία μεταβλητή X, έχουμε $X \to Y \to Z$.

Στη συνέχεια, αποδεικνύουμε ότι *η επεξεργασία της Y δεν μπορεί να αυξήσει το ποσό πληροφορίας που περιέχει η Y για την X.*

Θεώρημα: *Ανισότητα επεξεργασίας δεδομένων* (Data processing inequality). Αν $X \to Y \to Z$, τότε

$$I(X;Y) \geq I(X;Z). \tag{1.36}$$

Απόδειξη: Από τον κανόνα αλυσίδας για την αμοιβαία πληροφορία, μπορούμε να γράψουμε

$$\begin{aligned} I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \\ &= I(X;Y) + I(X;Z|Y) \end{aligned}$$

Αφού X και Z ανεξάρτητες δεδομένης της Y, δηλαδή, $I(X;Z|Y) = 0$, και $I(X;Y|Z) \geq 0$ (από τη μη αρνητικότητα της αμοιβαίας πληροφορίας), συμπεραίνουμε ότι $I(X;Y) \geq I(X;Z)$. □

Πόρισμα: Αν $Z = g(Y)$, τότε

$$I(X;Y) \geq I(X;g(Y)) \tag{1.37}$$

Απόδειξη: Η απόδειξη είναι προφανής αν θεωρήσουμε την αλυσία $X \to Y \to g(Y)$. □

Η ανισότητα (1.37) δηλώνει ότι *συναρτήσεις της Y δεν αυξάνουν την πληροφορία της Y για τη X.*

Πόρισμα: Αν $X \to Y \to Z$, τότε $I(X;Y|Z) \leq I(X;Y)$.

Απόδειξη: Από την (1.8), λαμβάνουμε

$$I(X;Y|Z) = -I(X;Z) + I(X;Y) + I(X;Z|Y).$$

Το πόρισμα αποδεικνύεται χρησιμοποιώντας ότι $I(X;Z|Y) = 0$ και $I(X;Z) \geq 0$. □

Η παραπάνω σχέση δηλώνει ότι αν $X \to Y \to Z$, τότε η εξάρτηση των X και Y μειώνεται (ή παραμένει αμετάβλητη) μετά την παρατήρηση της Z.

Κεφάλαιο 2

Ιδιότητα ασυμπτωτικής ισοδιαμέρισης

Η **ιδιότητα ασυμπτωτικής ισοδιαμέρισης** (asymptotic equipartition property) είναι απόρροια του ασθενούς νόμου των μεγάλων αριθμών που δηλώνει ότι αν $X_1, \ldots, X_n$ είναι ανεξάρτητες όμοια κατανεμημένες (α.ο.κ.) τ.μ. με συνάρτηση μάζας πιθανότητας $p_X(\cdot)$, τότε ο δειγματικός μέσος όρος $\frac{1}{n}\sum_{i=1}^{n} X_i$ είναι "κοντά" στο στατιστικό μέσο $E(X)$. Πιο συγκεκριμένα, αν $X_1, \ldots, X_n$ α.ο.κ. τ.μ. με μέση τιμή $\mu = E(X_i)$, τότε $\forall\, \epsilon > 0$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right) \to 0 \quad \text{όταν} \quad n \to \infty.$$

Η παραπάνω έκφραση γράφεται εναλλακτικά ως

$$-\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu \quad \text{όταν} \quad n \to \infty.$$

Θεώρημα: Αν $X_1, \ldots, X_n$ είναι α.ο.κ., με $X_i \sim p\,(x)$, τότε

$$-\frac{1}{n}\,\log p\,(X_1, \ldots, X_n) \xrightarrow{p} H(X).$$

Απόδειξη: Από την ανεξαρτησία των X_i λαμβάνουμε ότι $p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i)$. Οι τυχαίες μεταβλητές $\log p(X_i)$ είναι ανεξάρτητες, όμοια κατανεμημένες. Συνεπώς

$$\begin{aligned} -\frac{1}{n}\,\log p\,(X_1, \ldots, X_n) &= -\frac{1}{n}\,\log\left(\prod_{i=1}^{n} \log p(X_i)\right) \\ &= -\frac{1}{n}\sum_{i=1}^{n} \log p(X_i) \xrightarrow{p} -E\log p(X) = H(X). \end{aligned}$$

Ορισμός: Ορίζουμε σαν **τυπικό σύνολο** (typical set) $A_\epsilon^{(n)}$, ως προς τη συνάρτηση μάζας πιθανότητας $p(x)$, το σύνολο των ακολουθιών $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$ για τις οποίες ισχύει η σχέση

$$2^{-n(H(X)+\epsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\epsilon)} \tag{2.1}$$

με $p(x^n) = \prod_{i=1}^n p(x_i)$. Δηλαδή

$$A_\epsilon^{(n)} := \left\{ x^n \in \mathcal{X}^n : 2^{-n(H(X)+\epsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\epsilon)} \right\}. \tag{2.2}$$

Θεώρημα: *Ιδιότητες τυπικού συνόλου.*

1. Αν $x^n \in A_\epsilon^{(n)}$, τότε $H(X) - \epsilon \leq -\frac{1}{n} \log p(x^n) \leq H(X) + \epsilon$.
2. $\Pr\left[A_\epsilon^{(n)}\right] > 1 - \epsilon$, για αρκετά μεγάλο n.
3. $\left|A_\epsilon^{(n)}\right| \leq 2^{n(H(X)+\epsilon)}$, όπου $|A|$ είναι το πλήθος των στοιχείων του συνόλου A.
4. $\left|A_\epsilon^{(n)}\right| \geq (1-\epsilon)2^{n(H(X)-\epsilon)}$ για αρκούντως μεγάλο n.

Απόδειξη:

1. Άμεση από τον ορισμό και συνεπάγεται ότι $\left|-\frac{1}{n}\log p(x^n) - H(X)\right| < \epsilon$, για $x^n \in A_\epsilon^{(n)}$.
2. Από το νόμο των μεγάλων αριθμών, $\forall \epsilon > 0$, $\forall \delta > 0$, $\exists N_0$, ώστε $\forall n > N_0$

$$\Pr\left[x^n \in \mathcal{X}^n : \left|-\frac{1}{n}\log p(x^n) - H(X)\right| < \epsilon\right] > 1 - \delta.$$

   Θέτοντας $\delta = \epsilon$, λαμβάνουμε την ιδιότητα 2 (βεβαιωθείτε ότι αυτή η αντικατάσταση μπορεί να γίνει και για $\delta > \epsilon$ και για $\delta < \epsilon$).

3. $$1 = \sum_{x^n \in \mathcal{X}^n} p(x^n) \geq \sum_{x^n \in A_\epsilon^{(n)}} p(x^n) \geq \sum_{x^n \in A_\epsilon^{(n)}} 2^{-n(H(X)+\epsilon)} = 2^{-n(H(X)+\epsilon)} \left|A_\epsilon^{(n)}\right|.$$

4. Για αρκετά μεγάλο n, $\Pr\left[A_\epsilon^{(n)}\right] > 1 - \epsilon$, ώστε

$$1 - \epsilon < \Pr\left[A_\epsilon^{(n)}\right] = \sum_{x^n \in A_\epsilon^{(n)}} p(x^n) \leq \sum_{x^n \in A_\epsilon^{(n)}} 2^{-n(H(X)-\epsilon)} = 2^{-n(H(X)-\epsilon)} \left|A_\epsilon^{(n)}\right|.$$

Το Θεώρημα αποδείχθηκε. □

Η ποσότητα $-\frac{1}{n}\sum_{i=1}^{n}\log p(x_i)$ καλείται **εμπειρική εντροπία** (emperical entropy). Για τις τυπικές ακολουθίες, έχουμε ότι η εμπειρική εντροπία είναι "ϵ-κοντά" στη στατιστική εντροπία.

Συμπερασματικά, μπορούμε να πούμε ότι η πιθανότητα του τυπικού συνόλου $A_\epsilon^{(n)}$ είναι ≈ 1. Το πλήθος των στοιχείων του είναι $\approx 2^{nH(X)}$ και η πιθανότητα κάθε στοιχείου του $\approx 2^{-nH(X)}$. Οι παραπάνω ιδιότητες μπορούν να παραφραστούν ως εξής (ιδιότητα ασυμπτωτικής ισοδιαμέρισης): *"Σχεδόν όλα τα γεγονότα είναι σχεδόν ισοπίθανα."*

Παράδειγμα: Στην προσπάθεια να κατανοήσουμε καλύτερα τη δομή του τυπικού συνόλου, θεωρούμε την τυχαία μεταβλητή X με τιμές από το σύνολο $\{1,2,3\}$, με $p_X(1)=\frac{1}{2}$, $p_X(2)=\frac{1}{3}$ και $p_X(3)=\frac{1}{6}$ και εντροπία

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{3}\log\frac{1}{3} - \frac{1}{6}\log\frac{1}{6}.$$

Ας θεωρήσουμε την ακραία περίπτωση $\epsilon = 0$. Παρατηρήστε ότι για την ακολουθία μήκους n

$$x^n = \underbrace{1\ \ldots\ 1}_{\frac{n}{2}}\ \underbrace{2\ \ldots\ 2}_{\frac{n}{3}}\ \underbrace{3\ \ldots\ 3}_{\frac{n}{6}}$$

ισχύει

$$-\frac{1}{n}\log p(x^n) = -\frac{1}{n}\sum_{i=1}^{n}\log p(x_i) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{3}\log\frac{1}{3} - \frac{1}{6}\log\frac{1}{6}.$$

Συνεπώς, αυτή η ακολουθία είναι τυπική. Προσπαθήστε να καταλάβετε τι συμβαίνει με όλες τις μεταθέσεις της. Θεωρήστε ϵ μη μηδενικό αλλά μικρό. Πώς φαντάζεστε το τυπικό σύνολο σε αυτή την περίπτωση; Τι μπορείτε να πείτε για τη δειγματική πιθανότητα των $x_i \in \mathcal{X}$, $\frac{n_i}{n}$;

Προσπαθήστε να απαντήσετε στα παραπάνω ερωτήματα αν $p_X(1)=\frac{1}{2}$, $p_X(2)=\frac{1}{4}$ και $p_X(3)=\frac{1}{4}$. Ποιες ακολουθίες είναι τυπικές; Τι συμβαίνει με τη δειγματική πιθανότητα των στοιχείων 2 και 3 στις τυπικές ακολουθίες; □

Είδαμε ότι η πιθανότητα του τυπικού συνόλου είναι πολύ κοντά στη μονάδα. Ένα ενδιαφέρον ερώτημα είναι "*Πόσο μεγάλο είναι το πλήθος των στοιχείων του τυπικού συνόλου σε*

σχέση με το πλήθος όλων των δυνατών ακολουθιών $x^n \in \mathcal{X}^n$;" Για να απαντήσουμε στο ερώτημα θα πρέπει να υπολογίσουμε το λόγο $\frac{|A_\epsilon^{(n)}|}{|\mathcal{X}^n|}$. Αν $H(X) < \log|\mathcal{X}|$ (πότε δεν ισχύει αυτή η σχέση;), τότε

$$\frac{|A_\epsilon^{(n)}|}{|\mathcal{X}^n|} \approx \frac{2^{nH(X)}}{2^{n\log|\mathcal{X}|}} = 2^{-n(\log|\mathcal{X}|-H(X))} \to 0 \quad \text{όταν} \quad n \to \infty.$$

Συνεπώς, για αρκετά μεγάλο n, το πλήθος των στοιχείων του τυπικού συνόλου είναι εξαιρετικά μικρό σε σχέση με το πλήθος των δυνατών ακολουθιών, όμως η πιθανότητά του είναι πολύ κοντά στη μονάδα! Αυτή η σημαντική ιδιότητα καθιστά το τυπικό σύνολο ένα εξαιρετικό εργαλείο. Μία σημαντική εφαρμογή ακολουθεί.

2.1 Συμπίεση δεδομένων

Έστω $X_1, \ldots, X_n$ α.ο.κ. τ.μ. με κοινή συνάρτηση μάζας πιθανότητας $p(x)$. Επιθυμούμε να βρούμε βραχείες περιγραφές αυτών των τυχαίων ακολουθιών. Διαμερίζουμε όλες τις ακολουθίες του $\mathcal{X}^n$ σε δύο σύνολα: το τυπικό σύνολο $A_\epsilon^{(n)}$ και το συμπλήρωμά του $A_\epsilon^{(n)c}$.

Διατάσσουμε τις ακολουθίες κάθε συνόλου με κάποια διάταξη (έστω λεξικογραφική). Μπορούμε να αναπαραστήσουμε κάθε ακολουθία του $A_\epsilon^{(n)}$ με το δείκτη της ακολουθίας στη διάταξη. Αφού υπάρχουν $\leq 2^{n(H+\epsilon)}$ ακολουθίες στο $A_\epsilon^{(n)}$, η δεικτοδότηση απαιτεί $n(H+\epsilon)+1$ bits (το επιπλέον bit εισάγεται επειδή το $n(H+\epsilon)$ μπορεί να μην είναι ακέραιος). Μπροστά από κάθε τέτοια ακολουθία θέτουμε το 0 (για να δηλώσουμε το μήκος της κωδικής λέξης), δίνοντας συνολικό μήκος $n(H+\epsilon)+2$ bits σε καθεμία από τις παραπάνω κωδικές λέξεις.

Όμοια, μπορούμε να δεικτοδοτήσουμε κάθε ακολουθία στο συμπλήρωμα του $A_\epsilon^{(n)}$ με $n\log|\mathcal{X}|+1$ bits. Θέτοντας μπροστά από κάθε τέτοια ακολουθία το 1, έχουμε ένα κώδικα για όλες τις ακολουθίες του $\mathcal{X}^n$, ο οποίος είναι 1-1 και εύκολα αποκωδικοποιήσιμος. Για τον κώδικα αυτόν ισχύουν τα εξής:

1. Χρησιμοποιούμε $n\log|\mathcal{X}|+2$ bits για να περιγράψουμε τις μη-τυπικές ακολουθίες, παρόλο που χρειαζόμαστε (πολύ) λιγότερα. Παρόλο που αυτή η προσέγγιση δεν είναι βέλτιστη, όπως θα αποδείξουμε στη συνέχεια, είναι ικανή να οδηγήσει σε αποδοτική περιγραφή.

2. Οι τυπικές ακολουθίες έχουν βραχείες περιγραφές μήκους $\approx nH(X)$ bits.

Ορίζουμε ως $l(x^n)$ το μήκος της κωδικής λέξης που αντιστοιχεί στο $x^n = x_1 x_2 \cdots x_n$. Αν το n είναι αρκετά μεγάλο ώστε $\Pr\left[A_\epsilon^{(n)}\right] \geq 1-\epsilon$, τότε το **μέσο μήκος** (expected length) των κωδικών λέξεων είναι:

$$\begin{aligned} E\left(l(X^n)\right) &= \sum_{x\in\mathcal{X}^n} l(x^n)p(x^n) = \sum_{x^n\in A_\epsilon^{(n)}} l(x^n)p(x^n) + \sum_{x^n\in A_\epsilon^{(n)c}} l(x^n)p(x^n) \\ &= \sum_{x^n\in A_\epsilon^{(n)}} p(x^n)\left(n(H+\epsilon)+2\right) + \sum_{x^n\in A_\epsilon^{(n)c}} p(x^n)\left(n\log|\mathcal{X}|+2\right) \\ &= \left(n(H+\epsilon)+2\right)\Pr\left[A_\epsilon^{(n)}\right] + \left(n\log|\mathcal{X}|+2\right)\Pr\left[A_\epsilon^{(n)c}\right] \\ &\leq n(H+\epsilon) + \epsilon n\log|\mathcal{X}| + 2\epsilon + 2 = n(H+\epsilon') \end{aligned}$$

όπου $\epsilon' = \epsilon + \epsilon\log|\mathcal{X}| + \frac{2}{n} + \frac{2\epsilon}{n}$ μπορεί να γίνει αυθαίρετα μικρό με την κατάλληλη επιλογή των ϵ και n. Ανακεφαλαιώνοντας, έχουμε το ακόλουθο θεώρημα:

Θεώρημα: Έστω X^n ακολουθία α.ο.κ. $\sim p(x)$ και $\epsilon > 0$. Τότε, για n αρκετά μεγάλο, υπάρχει κώδικας που απεικονίζει μοναδικά (άρα ο κώδικας είναι αντιστρέψιμος) ακολουθίες μήκους n, x^n, σε δυαδικές ακολουθίες με μέσο μήκος δυαδικής ακολουθίας ανά σύμβολο

$$E\left(\frac{1}{n}l(X^n)\right) \leq H(X) + \epsilon. \tag{2.3}$$

2.2 Απο-κοινού τυπικές ακολουθίες

Έστω (X_i, Y_i), $i = 1, \ldots, n$, ανεξάρτητα όμοια κατανεμημένα ζεύγη τυχαίων μεταβλητών με από-κοινού συνάρτηση μάζας πιθανότητας $p(x, y)$.

Ορισμός 1 *Ορίζουμε ως σύνολο* **από-κοινού τυπικών** *(jointly typical) ακολουθιών, ως προς τη σμπ $p(x,y)$, το σύνολο $A_\epsilon^{(n)}$ που αποτελείται από τα ζεύγη (x^n, y^n) των οποίων οι εμπειρικές εντροπίες είναι ϵ-κοντά στις πραγματικές εντροπίες, δηλαδή*

$$\begin{aligned} A_\epsilon^{(n)} := \Bigg\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : &\left|-\frac{1}{n}\log p(x^n) - H(X)\right| < \epsilon, \\ &\left|-\frac{1}{n}\log p(y^n) - H(Y)\right| < \epsilon, \\ &\left|-\frac{1}{n}\log p(x^n,y^n) - H(X,Y)\right| < \epsilon\Bigg\} \end{aligned} \tag{2.4}$$

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

Θεώρημα 1 *Έστω (X^n, Y^n) ακολουθίες μήκους n με*

$$\Pr[X^n = x^n, Y^n = y^n] = p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i). \tag{2.5}$$

Τότε

1. $$\Pr\left((X^n, Y^n) \in A_\epsilon^{(n)}\right) \to 1 \quad \text{όταν} \quad n \to \infty. \tag{2.6}$$

2. $$\left|A_\epsilon^{(n)}\right| \leq 2^{n(H(X,Y)+\epsilon)}.$$

   Επίσης, για αρκετά μεγάλο n,

   $$|A_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H(X,Y)-\epsilon)}.$$

3. *Αν $\Pr(\tilde{X}^n = x^n, \tilde{Y}^n = y^n) = p(x^n)\,p(y^n) = \prod_{i=1}^{n} p(x_i)\,p(y_i)$, δηλαδή, $\tilde{X}^n$ και $\tilde{Y}^n$ ακολουθίες ανεξάρτητες μεταξύ τους με αοκ στοιχεία $\Pr(\tilde{X}_i = x) = p(x)$ και $\Pr(\tilde{Y}_i = y) = p(y)$, όπου οι $p(x)$ και $p(y)$ οι περιθώριες σμπ που προκύπτουν από την $p(x,y)$, τότε:*

   $$\Pr\left((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\right) \leq 2^{-n(I(X;Y)-3\epsilon)}. \tag{2.7}$$

   Επίσης, για αρκετά μεγάλο n

   $$\Pr\left((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\right) \geq (1-\epsilon)\,2^{-n(I(X;Y)+3\epsilon)}. \tag{2.8}$$

Απόδειξη

1. Από τον ασθενή νόμο των μεγάλων αριθμών

$$-\frac{1}{n}\log p(X^n) = -\frac{1}{n}\sum_{i=1}^{n}\log p(X_i) \xrightarrow{p} H(X).$$

Άρα, $\forall \epsilon > 0$, $\exists n_1$, ώστε $\forall n \geq n_1$ [1]

$$\Pr\left(\left|-\frac{1}{n}\log p(X^n) - H(X)\right| > \epsilon\right) < \frac{\epsilon}{3}. \tag{2.9}$$

[1] Μία αναλυτική έκφραση για την (2.9) είναι $\Pr\left(x^n \in \mathcal{X}^n : \left|-\frac{1}{n}\log p(x^n) - H(X)\right| > \epsilon\right) < \frac{\epsilon}{3}$.

Όμοια, $\forall \epsilon > 0$, $\exists n_2$, ώστε $\forall n \geq n_2$

$$\Pr\left(\left|-\frac{1}{n}\log p(Y^n) - H(Y)\right| > \epsilon\right) < \frac{\epsilon}{3}.$$

Επίσης

$$-\frac{1}{n}\log p(X^n, Y^n) = -\frac{1}{n}\sum_{i=1}^{n}\log p(X_i, Y_i) \xrightarrow{p} H(X, Y).$$

Άρα, $\forall \epsilon > 0$, $\exists n_3$, ώστε $\forall n \geq n_3$

$$\Pr\left(\left|-\frac{1}{n}\log p(X^n, Y^n) - H(X, Y)\right| > \epsilon\right) < \frac{\epsilon}{3}.$$

Χρησιμοποιώντας τις τρεις παραπάνω ανισότητες και το union bound, μπορεί να αποδειχθεί η 2.6.

2.

$$\begin{aligned}
1 &= \sum_{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n} p(x^n, y^n) \\
&\geq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} p(x^n, y^n) \\
&\geq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} 2^{-n(H(X,Y)+\epsilon)} \\
&= |A_\epsilon^{(n)}|\, 2^{-n(H(X,Y)+\epsilon)}
\end{aligned}$$

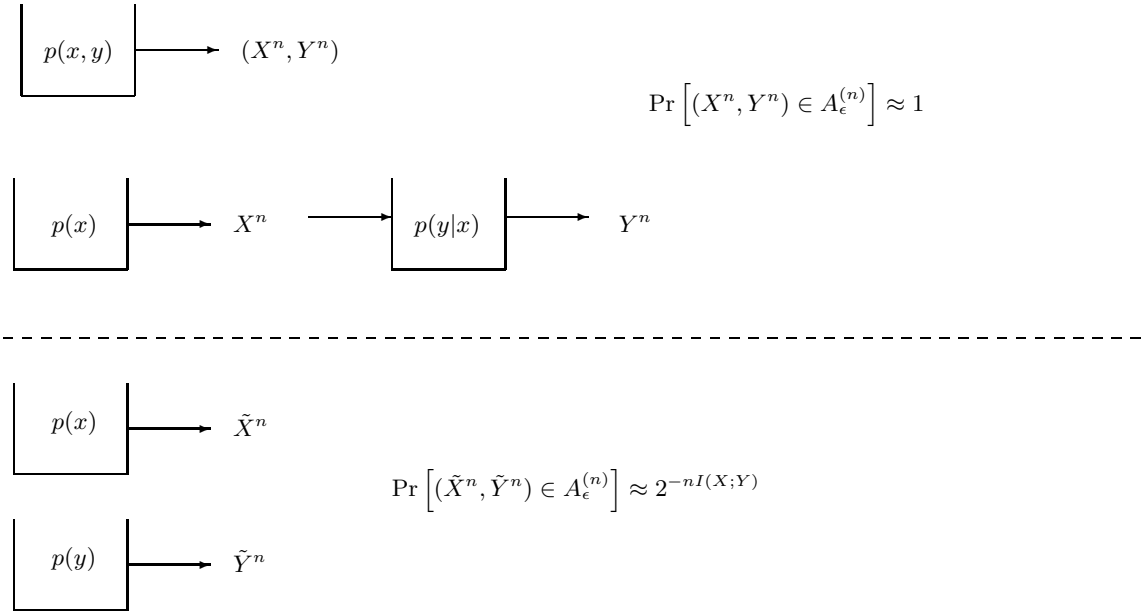
το οποίο αποδεικνύει την πρώτη ανισότητα του 2. Επιπλέον, για μεγάλο n, $\Pr\left(A_\epsilon^{(n)}\right) \geq 1-\epsilon$, και συνεπώς:

$$1 - \epsilon \leq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} p(x^n, y^n) \leq |A_\epsilon^{(n)}|\, 2^{-n(H(X,Y)-\epsilon)}$$

το οποίο αποδεικνύει τη δεύτερη ανισότητα του 2.

3. Αν $\tilde{X}^n, \tilde{Y}^n$ ανεξάρτητες με ίδιες περιθώριες όπως οι X^n και Y^n τότε:

$$\begin{aligned}
\Pr\left((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\right) &= \sum_{(x^n,\, y^n) \in A_\epsilon^{(n)}} p(x^n)\, p(y^n) \\
&\leq 2^{n(H(X,Y)+\epsilon)}\, 2^{-n(H(X)-\epsilon)}\, 2^{-n(H(Y)-\epsilon)} \\
&= 2^{-n\,(I(X;Y)-\,3\epsilon)}.
\end{aligned}$$



Σχήμα 2.1: Θεώρημα από-κοινού τυπικών ακολουθιών.

Όμοια:

$$\begin{aligned}\Pr\left((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\right) &= \sum_{(x^n, y^n)) \in A_\epsilon^{(n)}} p(x^n)\, p(y^n) \\ &\geq (1-\epsilon)\, 2^{n(H(X,Y)-\epsilon)}\, 2^{-n(H(X)+\epsilon)}\, 2^{-n(H(Y)+\epsilon)} \\ &= (1-\epsilon)\, 2^{-n\,(I(X;Y)+3\epsilon)}.\end{aligned}$$

Το Θεώρημα αποδείχθηκε. □

Στο Σχήμα 2.1 κατασκευάζουμε ζεύγη από-κοινού τυπικών ακολουθιών (X^n, Y^n) (1) επιλέγοντας ζεύγη $(X_i, Y_i) \sim p(x, y)$ και (2) επιλέγοντας αρχικά X^n με $X_i \sim p(x)$ και κατόπιν Y^n με $Y_i \sim p(y|x)$.[2] Από την άλλη, αν επιλέξουμε ανεξάρτητες $\tilde{X}^n$ και $\tilde{Y}^n$ με $\tilde{X}_i \sim p(x)$ και $\tilde{Y}_i \sim p(y)$, τότε το ζεύγος $(\tilde{X}^n, \tilde{Y}^n)$ είναι από-κοινού τυπικό με πιθανότητα $\approx 2^{-nI(X;Y)}$.

Σχόλια: Υπάρχουν $\approx 2^{nH(X)}$ τυπικές ακολουθίες X^n, $\approx 2^{nH(Y)}$ τυπικές Y^n και *μόνο* $\approx 2^{nH(X,Y)}$ από-κοινού τυπικές (X^n, Y^n) (θυμηθείτε ότι $H(X,Y) \leq H(X) + H(Y)$). Αν κατασκευάσουμε ζεύγος ακολουθιών (X^n, Y^n) επιλέγοντας τα ζεύγη (X_i, Y_i) ανεξάρτητα όμοια

[2] Όπως θα δούμε στη συνέχεια, ο δεύτερος τρόπος εγγυάται ότι η πραγματική είσοδος και η έξοδος του καναλιού είναι από-κοινού τυπικές με πιθανότητα ≈ 1.

κατανεμημένα με από-κοινού συνάρτηση μάζας πιθανότητας $p(x, y)$, τότε το ζεύγος (X^n, Y^n) είναι τυπικό με πιθανότητα ≈ 1.

Αν, όμως, κατασκευάσουμε ακολουθίες $\tilde{X}^n$ και $\tilde{Y}^n$ ανεξάρτητες μεταξύ τους με ανεξάρτητα όμοια κατανεμημένα στοιχεία X_i και Y_i με συναρτήσεις μάζας πιθανότητας $p(x)$ και $p(y)$, αντίστοιχα, τότε καθεμία από τις $\tilde{X}^n$ και $\tilde{Y}^n$ είναι τυπική, ως προς την $p(x)$ και $p(y)$, αντίστοιχα, με πιθανότητα ≈ 1, όμως το ζεύγος $(\tilde{X}^n, \tilde{Y}^n)$ είναι από-κοινού τυπικό, ως προς την $p(x, y)$, με πιθανότητα $\approx 2^{-nI(X;Y)}$ (Πότε η πιθανότητα αυτή είναι ≈ 1; Όταν δεν συμβαίνει αυτό παρατηρήστε ότι η πιθανότητα τείνει στο 0 εκθετικά!). Άρα, για δεδομένη Y^n, κάποιος πρέπει να θεωρήσει $\approx 2^{n\,I(X;Y)}$ ανεξάρτητες ακολουθίες X^n πριν να είναι πολύ πιθανό να σχηματίσει ένα από-κοινού τυπικό ζεύγος.

Οι παραπάνω ιδιότητας είναι εξαιρετικά χρήσιμες για την απόδειξη και την κατανόηση των βασικών θεωρημάτων της Θεωρίας Πληροφορίας. □

Κεφάλαιο 3

Χωρητικότητα διακριτού καναλιού χωρίς μνήμη

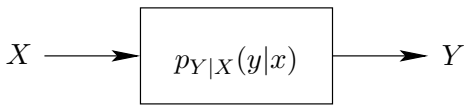
Σε αυτό το κεφάλαιο, θα αποδείξουμε το πρώτο σημαντικό αποτέλεσμα της Θεωρίας Πληροφορίας, το θεώρημα χωρητικότητας διακριτού καναλιού χωρίς μνήμη του Shannon.

3.1 Διακριτό κανάλι - Χωρητικότητα

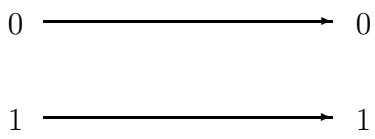
Ορισμός 2 *Το σύστημα $(\mathcal{X}, p(y|x), \mathcal{Y})$, όπου $\mathcal{X}$ είναι το αλφάβητο εισόδου, $\mathcal{Y}$ το αλφάβητο εξόδου και $p(y|x)$ η "συνάρτηση μεταφοράς" του συστήματος (δηλαδή, η πιθανότητα να παρατηρήσουμε έξοδο y, δεδομένου ότι η είσοδος είναι x), καλείται* **διακριτό κανάλι** *(discrete channel).* ◇

Ορισμός 3 *Ένα κανάλι καλείται* **κανάλι χωρίς μνήμη** *(memoryless channel) αν η έξοδος μία χρονική στιγμή εξαρτάται* **μόνο** *από την είσοδο την ίδια χρονική στιγμή και είναι υπό-συνθήκη ανεξάρτητη από προηγούμενες εισόδους–εξόδους.* ◇

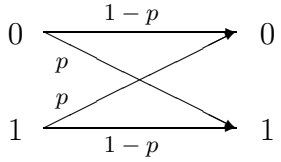
Ορισμός 4 *Έστω διακριτό κανάλι χωρίς μνήμη, με είσοδο X, έξοδο Y και συνάρτηση*



Σχήμα 3.1: Διακριτό κανάλι χωρίς μνήμη.



Σχήμα 3.2: Αθόρυβο δυαδικό κανάλι.



Σχήμα 3.3: Δυαδικό συμμετρικό κανάλι.

μεταφοράς $p(y|x)$. *Η ποσότητα*

$$C := \max_{p(x)} I(X;Y), \tag{3.1}$$

όπου η μεγιστοποίηση πραγματοποιείται πάνω σε όλες τις δυνατές συναρτήσεις μάζας πιθανότητας της εισόδου, $p(x)$, *καλείται* **χωρητικότητα πληροφορίας** *(information capacity) του καναλιού.* ◇

Θα αποδείξουμε ότι η χωρητικότητα πληροφορίας, C, ισούται με τη **λειτουργική χωρητικότητα** (operational capacity) του καναλιού, δηλαδή, με τον υψηλότερο ρυθμό bits/(χρήση καναλιού) με τον οποίο μπορούμε να μεταδώσουμε πληροφορία μέσω του καναλιού με αυθαίρετα μικρή πιθανότητα σφάλματος.

Παρατήρηση: Όπως έχουμε αποδείξει, για δεδομένη $p_{Y|X}(y|x)$, η αμοιβαία πληροφορία $I(X;Y)$ είναι κοίλη συνάρτηση της σμπ $p(x)$. Χρησιμοποιώντας επιχειρήματα συνέχειας της $I(X;Y)$, ως προς την $p(x)$, και συμπάγειας και κυρτότητας του συνόλου των δυνατών $p(x)$, μπορεί να αποδειχθεί ότι η ποσότητα $\max_{p(x)} I(X;Y)$ υπάρχει. Η χωρητικότητα μπορεί να εκφραστεί σε κλειστή μορφή σε πολύ λίγες, αλλά σημαντικές, περιπτώσεις. Γενικά, υπολογίζεται αριθμητικά μέσω επαναληπτικής αριθμητικής διαδικασίας.

Παράδειγμα 1 Αθόρυβο Δυαδικό Κανάλι *(Binary Noiseless Channel). Η χωρητικότητα του καναλιού του Σχήματος 3.2 είναι* $C = 1$ *bit ανά χρήση καναλιού.*

Απόδειξη:

$$\begin{aligned} C &= \max_{p(x)}(H(X) - H(X|Y)) = \max_{p(x)} H(x) \\ &= \max_{p}(-p\,\log p - (1-p)\,\log(1-p)) = 1. \end{aligned}$$

Το μέγιστο επιτυγχάνεται για $p = \frac{1}{2}$. □

Παράδειγμα 2 Δυαδικό Συμμετρικό Κανάλι *(Binary Symmetric Channel): Η χωρητικότητα του καναλιού του Σχήματος 3.3 είναι*

$$C_{\text{BSC}} = 1 - H(p). \tag{3.2}$$

Απόδειξη:

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum p(x)\, H(Y|X = x). \end{aligned} \tag{3.3}$$

Από τον ορισμό της υπό-συνθήκη εντροπίας της Y, δεδομένου ότι $X = x$, έχουμε

$$\begin{aligned} H(Y|X=0) &= -\sum_{y=0,1} p_{Y|X}(y|0) \log p_{Y|X}(y|0) \\ &= -p_{Y|X}(0|0) \log p_{Y|X}(0|0) - p_{Y|X}(1|0) \log p_{Y|X}(1|0) \\ &= -(1-p)\log(1-p) - p\log p \\ &= H(p). \end{aligned} \tag{3.4}$$

Μπορεί να αποδειχθεί εύκολα (να το αποδείξετε) ότι $H(Y|X=1) = H(p)$. Συνεπώς,

$$I(X;Y) = H(Y) - H(p) \leq 1 - H(p). \tag{3.5}$$

Η ισότητα στη σχέση (3.5) επιτυγχάνεται για $p_X(0) = p_X(1) = \frac{1}{2}$ (να το αποδείξετε). □

Ιδιότητες της χωρητικότητας

1. $C \geq 0$, διότι $I(X;Y) \geq 0$.

2. $C \leq \min\{\log|\mathcal{X}|, \log|\mathcal{Y}|\}$ διότι, για κάθε $p(x)$,

$$I(X;Y) = H(X) - H(X|Y) \leq H(X) \leq \log|\mathcal{X}|$$

και άρα $C = \max_{p(x)} I(X;Y) \leq \log|\mathcal{X}|$. Ομοίως, $C \leq \log|\mathcal{Y}|$.

Ορισμός 5 *Η n-οστή επέκταση ενός διακριτού καναλιού χωρίς μνήμη είναι το κανάλι*

$$(\mathcal{X}^n, p(y^n|x^n), \mathcal{Y}^n),$$

όπου $p(y_k|x^k, y^{k-1}) = p(y_k|x_k)$, για $k = 1, \ldots, n$. ◇

Παρατήρηση: Αν το κανάλι χρησιμοποιείται χωρίς ανάδραση, δηλαδή αν η είσοδος είναι ανεξάρτητη παρελθόντων εξόδων, δηλαδή,

$$p(x_k|x^{k-1}, y^{k-1}) = p(x_k|x^{k-1}) \tag{3.6}$$

τότε:

$$p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i). \tag{3.7}$$

Ορισμός 6 *Ένας (M, n) κώδικας για το κανάλι $(\mathcal{X}, p(y|x), \mathcal{Y})$ αποτελείται από τα εξής:*

1. *Ένα σύνολο δεικτών $\mathcal{W} := \{1, 2, \ldots, M\}$, το οποίο αναπαριστά τα M δυνατά μηνύματα.*

2. *Μία συνάρτηση κωδικοποίησης*

$$X^n : \mathcal{W} \to \mathcal{X}^n$$

η οποία παράγει τις κωδικές λέξεις $X^n(1), \ldots, X^n(M)$. Το σύνολο των κωδικών λέξεων καλείται **κωδικό βιβλίο** *(codebook). Όταν θέλουμε να μεταδώσουμε το μήνυμα i θέτουμε ως είσοδο στο κανάλι την κωδική λέξη $X^n(i)$.*

3. *Μία συνάρτηση αποκωδικοποίησης*

$$g : \mathcal{Y}^n \to \mathcal{W}$$

η οποία είναι ένας ντετερμινιστικός κανόνας που παράγει μία εκτίμηση για το μήνυμα που μεταδόθηκε για κάθε ακολουθία εξόδου.

◇

Ορισμός 7 *Πιθανότητα σφάλματος δεδομένου ότι μεταδόθηκε το i-οστό μήνυμα συγκεκριμένου κώδικα:*

$$\lambda_i := \Pr(g(Y^n) \neq i | X^n = x^n(i)) = \sum_{y^n \in \mathcal{Y}^n} p(y^n|x^n(i))\, I(g(y^n) \neq i) \tag{3.8}$$

όπου $I(\cdot)$ είναι η συνάρτηση δείκτης

$$I(P) := \begin{cases} 1, & \text{αν } P \text{ αληθής πρόταση,} \\ 0, & \text{αν } P \text{ ψευδής πρόταση.} \end{cases}$$

◇

Ορισμός 8 *Μέγιστη πιθανότητα σφάλματος $\lambda^{(n)}$ ενός συγκεκριμένου (M, n) κώδικα:*

$$\lambda^{(n)} := \max_{i \in \{1,\ldots,M\}} \lambda_i. \tag{3.9}$$

◇

Ορισμός 9 *Αριθμητικός μέσος πιθανότητας σφάλματος ενός συγκεκριμένου (M, n) κώδικα:*

$$P_e^{(n)} := \frac{1}{M} \sum_{i=1}^{M} \lambda_i. \tag{3.10}$$

◇

Αν ο δείκτης W επιλέγεται με βάση την ομοιόμορφη κατανομή, τότε η $P_e^{(n)}$ είναι η μέση πιθανότητα σφάλματος του κώδικα. Προφανώς $P_e^{(n)} \leq \lambda^{(n)}$.

Ορισμός 10 *Ο ρυθμός R ενός (M, n) κώδικα είναι*

$$R := \frac{\log M}{n} \text{ bits ανά χρήση καναλιού.} \tag{3.11}$$

◇

Ορισμός 11 *Ο ρυθμός R καλείται* επιτεύξιμος *αν υπάρχει ακολουθία $(\lceil 2^{nR} \rceil, n)$ κωδίκων τέτοια ώστε η μέγιστη πιθανότητα σφάλματος $\lambda^{(n)}$ να τείνει στο 0 όταν $n \to \infty$.* ◇

Παρατήρηση: Για απλούστευση συμβολισμού, θα γράφουμε $(2^{nR}, n)$ αντί για $(\lceil 2^{nR} \rceil, n)$.

Ορισμός 12 *Το ελάχιστο άνω φράγμα (supremum) των επιτεύξιμων ρυθμών, καλείται λειτουργική χωρητικότητα, ή, απλά, χωρητικότητα, ενός διακριτού καναλιού χωρίς μνήμη.* ◇

3.2 Θεώρημα κωδικοποίησης

Θεώρημα 2 Θεώρημα κωδικοποίησης Shannon. *Για κάθε ρυθμό $R < C$, υπάρχει ακολουθία $(2^{nR}, n)$ κωδίκων με μέγιστη πιθανότητα σφάλματος $\lambda^{(n)} \to 0$ όταν $n \to \infty$. Αντίστροφα, για κάθε ακολουθία $(2^{nR}, n)$ κωδίκων με $\lambda^{(n)} \to 0$, θα πρέπει να ισχύει $R \leq C$.*

Άρα, για ρυθμούς μικρότερους της χωρητικότητας επιτυγχάνουμε αυθαίρετα μικρή πιθανότητα σφάλματος, για αρκετά μεγάλο n. Το αποτέλεσμα είναι ενάντια στη διαίσθηση. Αφού το κανάλι εισάγει σφάλματα, κάθε προσπάθεια "διόρθωσης" υπόκειται σε σφάλματα και το ίδιο επαναλαμβάνεται επ΄ άπειρο. Η κρατούσα άποψη μέχρι τον Shannon ήταν ότι, για να επιτύχουμε αυθαίρετα μικρή πιθανότητα σφάλματος, θα πρέπει να μειώσουμε αντίστοιχα το ρυθμό μετάδοσης πληροφορίας. Ο Shannon απέδειξε ότι αρκεί $R < C$.

Η απόδειξη του θεωρήματος βασίζεται στις ιδιότητες των από-κοινού τυπικών ακολουθιών.

Το θεώρημα με λόγια: Η απόδειξη του Shannon βασίζεται στο λεγόμενο random coding argument. Ουσιαστικά, αποδεικνύεται ότι, για κάθε $\epsilon > 0$, αν $R < C$ και n αρκετά μεγάλο, τότε η μέση πιθανότητα σφάλματος πάνω σε όλους τους κώδικες και όλες τις εξόδους είναι $\leq 2\epsilon$. Συνεπώς, υπάρχει κάποιος κώδικας με μέση πιθανότητα σφάλματος $P_e^{(n)} \leq 2\epsilon$. Με ένα απλό επιχείρημα μπορεί να αποδειχθεί ότι υπάρχει κώδικας με μέγιστη πιθανότητα σφάλματος $\lambda^{(n)} \leq 4\epsilon$. Αφού το ϵ είναι αυθαίρετο, συμπεραίνουμε ότι, αν $R < C$ και n αρκετά μεγάλο, τότε υπάρχει κώδικας που επιτυγχάνει αυθαίρετα καλή ποιότητα επικοινωνίας. □

3.2.1 Απόδειξη επιτευξιμότητας

Σε αυτό το εδάφιο, θα αποδείξουμε ότι *αν $R < C$, τότε υπάρχει ακολουθία $(2^{nR}, n)$ κωδίκων με $\lambda^{(n)} \to 0$, όταν $n \to \infty$.*

Θεωρούμε την εξής ακολουθία γεγονότων:

1. Επιλέγουμε συνάρτηση μάζας πιθανότητας $p(x)$. Κατασκευάζουμε έναν $(2^{nR}, n)$ κώδικα τυχαία δημιουργώντας 2^{nR} ανεξάρτητες κωδικές λέξεις, μήκους n η καθεμία, με τα στοιχεία κάθε κωδικής λέξης ανεξάρτητα, όμοια κατανεμημένα, με συνάρτηση μάζας πιθανότητας $p(x)$. Οι τυχαίες κωδικές λέξεις συμβολίζονται ως $X^n(1), \ldots, X^n(2^{nR})$. Ένας τυχαίος κώδικας $\mathcal{C}$ αποτελείται από τις τυχαίες κωδικές λέξεις $\mathcal{C} = \left[X^n(1), \ldots, X^n\left(2^{nR}\right)\right]$. Ένας συγκεκριμένος κώδικας συμβολίζεται ως $\mathtt{C} = \left[x^n(1), \ldots, x^n\left(2^{nR}\right)\right]$.

2. Φανερώνουμε τον κώδικα $\mathcal{C}$ στον πομπό και στο δέκτη, ο οποίος γνωρίζει επιπλέον την $p(y|x)$.

3. Επιλέγουμε το μήνυμα W ομοιόμορφα,[1] δηλαδή, $\Pr(W = w) = 2^{-nR}$, για $w = 1, 2, \ldots, 2^{nR}$.

4. Η κωδική λέξη $X^n(W)$ τίθεται σαν είσοδος στο κανάλι.

5. Ο δέκτης λαμβάνει Y^n με $P(Y^n = y^n | X^n(W) = x^n) = \prod_{i=1}^{n} p(y_i|x_i)$.

6. Ο δέκτης αποφασίζει ποιο μήνυμα στάλθηκε χρησιμοποιώντας αποκωδικοποίηση από-κοινού τυπικότητας (jointly typical decoding).[2] Αυτό σημαίνει ότι ο δέκτης αποφασίζει ότι το μήνυμα που εστάλη είναι το $\hat{W}$ αν το ζεύγος $(X^n(\hat{W}), Y^n)$ είναι από-κοινού τυπικό. Αν δεν υπάρχει τέτοιο $\hat{W}$ ή αν υπάρχουν περισσότερα από ένα, τότε δηλώνεται σφάλμα.

Περιγραφή απόδειξης: Θα υπολογίσουμε τη μέση πιθανότητα σφάλματος πάνω σε όλους τους δυνατούς κώδικες οι οποίοι κατασκευάζονται με βάση την $p(x)$. Υπάρχουν δύο πηγές σφάλματος αποκωδικοποίησης:

1. Η είσοδος και η έξοδος δεν είναι από-κοινού τυπικές. Αυτό το γεγονός συμβαίνει με πιθανότητα η οποία μπορεί να γίνει αυθαίρετα μικρή για αρκετά μεγάλο n.

2. Περισσότερες από μία είσοδοι είναι από-κοινού τυπικές με την έξοδο. Για είσοδο εκτός της πραγματικής, η πιθανότητα να είναι από κοινού τυπική με την έξοδο είναι $\approx 2^{-nI(X;Y)}$. Μπορούμε να θεωρήσουμε 2^{nR} κωδικές λέξεις με $R < I(X;Y)$, και να φράξουμε την πιθανότητα σφάλματος από μία ποσότητα $\approx 2^{n(R-I(X;Y))}$, η οποία μπορεί να γίνει αυθαίρετα μικρή αν το n επιλεγεί αρκετά μεγάλο.

Απόδειξη: Η απόδειξη ακολουθεί τα βήματα του McEliece. Θα κάνουμε εκτενή χρήση του θεωρήματος ολικής πιθανότητας στη μορφή

$$P(A|Y = y) = \sum_{x} p_{X|Y}(x|y) P(A|X = x, Y = y). \tag{3.12}$$

[1] Μπορεί να αποδειχθεί ότι το θεώρημα ισχύει για οποιαδήποτε κατανομή του W.

[2] Η αποκωδικοποίηση μέσω τυπικών ακολουθιών είναι υποβέλτιστη. Όμως είναι απλή στην ανάλυση και επιτυγχάνει όλους τους ρυθμούς μέχρι τη χωρητικότητα! Βέλτιση αποκωδικοποίηση επιτυγχάνεται μέσω της τεχνικής της μέγιστης πιθανοφάνειας (maximum likelihood).

1: Έστω *δεδομένος* κώδικας $\mathtt{C} = [x^n(1), \ldots, x^n(2^{nR})]$, *δεδομένος* δείκτης μηνύματος $W = i$, και *δεδομένη* έξοδος $Y^n = y^n$. Τότε, θεωρώντας αποκωδικοποίηση με χρήση από-κοινού τυπικών ακολουθιών, έχουμε

$$\begin{aligned} & P\left(e \,|\, W = i, \mathcal{C} = \mathtt{C}, Y^n = y^n\right) \\ & \quad = P\left((x^n(i), y^n) \notin A_\epsilon^{(n)} \cup \bigcup_{\substack{j=1 \\ j \neq i}}^{2^{nR}} (x^n(j), y^n) \in A_\epsilon^{(n)} \right) \\ & \quad \overset{(a)}{\leq} P\left((x^n(i), y^n) \notin A_\epsilon^{(n)} \right) + \sum_{\substack{j=1 \\ j \neq i}}^{2^{nR}} P\left((x^n(j), y^n) \in A_\epsilon^{(n)} \right) \end{aligned}$$

όπου στο σημείο (a) χρησιμοποιήσαμε το φράγμα ένωσης. Ορίζουμε

$$\begin{aligned} \phi(x^n, y^n) &:= \begin{cases} 1, & \text{αν } (x^n, y^n) \in A_\epsilon^{(n)} \\ 0, & \text{διαφορετικά} \end{cases}, \\ \phi^c(x^n, y^n) &:= 1 - \phi(x^n, y^n). \end{aligned} \tag{3.13}$$

Τότε

$$P\left(e \,|\, W = i, \mathcal{C} = \mathtt{C}, Y^n = y^n\right) \leq \phi^c(x^n(i), y^n) + \sum_{\substack{j=1 \\ j \neq i}}^{2^{nR}} \phi(x^n(j), y^n). \tag{3.14}$$

Παίρνοντας μέση τιμή, πάνω σε όλες τις εξόδους, δεδομένου ότι η είσοδος είναι η $x^n(i)$, λαμβάνουμε

$$\begin{aligned} & P\left(e \,|\, W = i, \mathcal{C} = \mathtt{C}\right) \\ & \quad \leq \sum_{y^n \in \mathcal{Y}^n} \phi^c(x^n(i), y^n)\, p(y^n | x^n(i)) + \sum_{\substack{j=1 \\ j \neq i}}^{2^{nR}} \sum_{y^n \in \mathcal{Y}^n} \phi(x^n(j), y^n)\, p(y^n | x^n(i)) \\ & \quad = Q_i(x^n(1), \ldots, x^n(2^{nR})). \end{aligned} \tag{3.15}$$

Θέλουμε να κατασκευάσουμε κώδικα $[x^n(1), \ldots, x^n(2^{nR})]$ για τον οποίο η ποσότητα Q_i να είναι ταυτόχρονα μικρή για όλα τα i. Όμως, ο υπολογισμός του Q_i είναι, γενικά, εξαιρετικά δύσκολος. Πιο εύκολος είναι ο υπολογισμός της μέσης τιμής του Q_i, πάνω σε όλους τους δυνατούς κώδικες.

2: Η πιθανότητα του κώδικα $\mathtt{C} = [x^n(1), \ldots, x^n(2^{nR})]$ ισούται με

$$P(\mathtt{C}) = p(x^n(1), \ldots, x^n(2^{nR})) = \prod_{i=1}^{2^{nR}} p(x^n(i)) \tag{3.16}$$

όπου, αν $x^n(i) = (x(i,1), \ldots, x(i,n))$, τότε $p(x^n(i)) = \prod_{j=1}^{n} p(x(i,j))$.

Στη συνέχεια, θεωρούμε την ποσότητα $Q_i(x^n(1), \ldots, x^n(2^{nR}))$ σαν τυχαία μεταβλητή. Παίρνοντας μέση τιμή, πάνω σε όλους τους κώδικες, λαμβάνουμε

$$\begin{aligned} P(E|W=i) &\leq \mathcal{E}(Q_i) \\ &= \sum_{(x^n(1),\ldots,x^n(2^{nR}))\in(\mathcal{X}^n)^{2^{nR}}} p(x^n(1)) \cdots p(x^n(2^{nR}))\, Q_i(x^n(1), \ldots, x^n(2^{nR})). \end{aligned} \tag{3.17}$$

Υπολογίζουμε τους όρους

$$\begin{aligned} E_1 &= \sum_{(x^n(1),\ldots,x^n(2^{nR}))\in(\mathcal{X}^n)^{2^{nR}}} p(x^n(1)) \cdots p(x^n(2^{nR})) \sum_{y^n\in\mathcal{Y}^n} \phi^c(x^n(i), y^n)\, p(y^n|x^n(i)) \\ &\stackrel{!}{=} \sum_{(x^n(i),y^n)\in\mathcal{X}^n\times\mathcal{Y}^n} p(x^n(i))\, \phi^c(x^n(i), y^n)\, p(y^n|x^n(i)) \\ &= \sum_{(x^n(i),y^n)\in\mathcal{X}^n\times\mathcal{Y}^n} p(x^n(i), y^n)\, \phi^c(x^n(i), y^n) \\ &= \sum_{(x^n,y^n)\notin A_\epsilon^{(n)}} p(x^n, y^n) \end{aligned} \tag{3.18}$$

και

$$\begin{aligned} E_2^j &= \sum_{(x^n(1),\ldots,x^n(2^{nR}))\in(\mathcal{X}^n)^{2^{nR}}} p(x^n(1)) \cdots p(x^n(2^{nR})) \sum_{y^n\in\mathcal{Y}^n} \phi(x^n(j), y^n)\, p(y^n|x^n(i)) \\ &\stackrel{!}{=} \sum_{(x^n(j),y^n)\in\mathcal{X}^n\times\mathcal{Y}^n} p(x^n(j))\phi(x^n(j), y^n) \sum_{x^n(i)\in\mathcal{X}^n} p(x^n(i))p(y^n|x^n(i)) \\ &= \sum_{(x^n(j),y^n)\in\mathcal{X}^n\times\mathcal{Y}^n} p(x^n(j))\phi(x^n(j), y^n)p(y^n) \\ &= \sum_{(x^n,y^n)\in A_\epsilon^{(n)}} p(x^n)\, p(y^n). \end{aligned} \tag{3.19}$$

Για n αρκετά μεγάλο, έχουμε ότι $E_1 < \epsilon$. Επιπλέον, για κάθε $j = 1, \ldots, 2^{nR}$, με $j \neq i$, έχουμε ότι $E_2^j \leq 2^{-n(I(X;Y)-3\epsilon)}$.

Συνεπώς

$$P(e|W=i) \leq \mathcal{E}(Q_i) \leq \epsilon + 2^{nR}2^{-n(I(X;Y)-3\epsilon)} \leq 2\epsilon \tag{3.20}$$

αν $R < I(X;Y) - 3\epsilon$ και n αρκετά μεγάλο.

3: Παρατηρούμε ότι το εξαχθέν άνω φράγμα είναι ανεξάρτητο της τιμής του W. Αν οι δείκτες μηνύματος W επιλέγονται ομοιόμορφα,[3] τότε

$$P(e) = \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} P(e|W=i) \leq \epsilon + 2^{nR}\, 2^{-n(I(X;Y)-3\epsilon)} \leq 2\epsilon \tag{3.21}$$

αν $R < I(X;Y) - 3\epsilon$ και n αρκετά μεγάλο.

Επίλογος

1. Επιλέγουμε $p^*(x)$ το οποίο επιτυγχάνει τη χωρητικότητα C. Τότε, η συνθήκη $R < I(X;Y) - 3\epsilon$ μετατρέπεται σε $R < C - 3\epsilon$.

2. Αφού η μέση πιθανότητα σφάλματος πάνω σε όλους τους κώδικες είναι $\leq 2\epsilon$, υπάρχει κώδικας $\mathcal{C}^*$ με μέση πιθανότητα σφάλματος $P_e^{(n)}(\mathcal{C}^*) \leq 2\epsilon$.

3. Αν αποκλείσουμε τις χειρότερες μισές κωδικές λέξεις του κώδικα $\mathcal{C}^*$, όπου η απόδοση κάθε κωδικής λέξης κρίνεται με βάση την (3.8), τότε προκύπτει κώδικας με μέγιστη πιθανότητα σφάλματος $\lambda^{(n)} \leq 4\epsilon$ (διαφορετικά, αποδείξτε ότι $P_e^{(n)}(\mathcal{C}^*) > 2\epsilon$, το οποίο είναι άτοπο) και ρυθμό $R - \frac{1}{n}$ (η μείωση του ρυθμού είναι ασήμαντη για μεγάλα n).

Έτσι, κατασκευάστηκε κώδικας με ρυθμό $R' = R - \dfrac{1}{n}$ και μέγιστη πιθανότητα σφάλματος $\lambda^{(n)} \leq 4\epsilon$. Αφού το ϵ είναι αυθαίρετο, τα παραπάνω αποδεικνύουν την επιτευξιμότητα κάθε ρυθμού μετάδοσης μέχρι τη χωρητικότητα. □

Η μέθοδος "τυχαίας κωδικοποίησης" (random coding) λύνει ένα πολύ δύσκολο πρόβλημα, αποδεικνύοντας την ύπαρξη ενός καλού κώδικα. Για να βρούμε έναν καλό κώδικα, θεωρητικά, θα μπορούσαμε να ελέγξουμε όλους τους δυνατούς κώδικες μήκους n, το πλήθος των οποίων είναι $|\mathcal{X}|^{n2^{nR}}$. Για $2^{nR} \approx 2^{nC}$, το πλήθος των κωδίκων είναι $\approx |\mathcal{X}|^{n2^{nC}}$, το οποίο είναι εξαιρετικά μεγάλο για μεγάλο n.

[3] Προφανώς, η ανισότητα $P(e) \leq 2\epsilon$ ισχύει για οποιαδήποτε κατανομή του W.

Στη συνέχεια, αποδεικνύουμε ότι αν κατασκευάσουμε έναν τυχαίο κώδικα με $R < C$, τότε, για αρκετά μεγάλο n, αυτός ο κώδικας θα είναι "καλός" με μεγάλη πιθανότητα. Όπως αποδείξαμε, αν $R < C$, τότε $P(e) \leq 2\epsilon$, για αρκετά μεγάλο n. Τότε, μπορούμε να γράψουμε

$$\begin{aligned} P(e) &= \sum_{\mathtt{C}} P(\mathtt{C})\, P(e|\mathtt{C}) \\ &= \sum_{\mathtt{C}:P(e|\mathtt{C})\leq\psi} P(\mathtt{C})\, P(e|\mathtt{C}) + \sum_{\mathtt{C}:P(e|\mathtt{C})>\psi} P(\mathtt{C})\, P(e|\mathtt{C}) \\ &\geq \sum_{\mathtt{C}:P(e|\mathtt{C})>\psi} P(\mathtt{C})\, P(e|\mathtt{C}) > \psi \sum_{\mathtt{C}:P(e|\mathtt{C})>\psi} P(\mathtt{C}). \end{aligned} \tag{3.22}$$

Συνεπώς, η πιθανότητα του συνόλου των "κακών" κωδίκων φράσσεται από πάνω ως εξής:

$$\sum_{\mathtt{C}:P(e|\mathtt{C})>\psi} P(\mathtt{C}) < \frac{P(e)}{\psi} \leq \frac{2\epsilon}{\psi}. \tag{3.23}$$

Για δεδομένο ψ, αυτό το άνω φράγμα μπορεί να γίνει όσο μικρό θέλουμε αν επιλέξουμε κατάλληλα το n (διότι, αν $R < C$, το ϵ μπορεί να γίνει αυθαίρετα μικρό για αρκούντως μεγάλο n). Αυτό σημαίνει ότι, για κάθε ψ, η πιθανότητα του συνόλου των κωδίκων με μέση πιθανότητα σφάλματος $> \psi$ μπορεί να γίνει αυθαίρετα μικρή για αρκετά μεγάλο n.

Το σημαντικό μειονέκτημα αυτής της μορφής κωδικοποίησης είναι η τεράστια υπολογιστική πολυπλοκότητα που απαιτείται για την αποκωδικοποίηση, λόγω έλλειψης δομής. Το γεγονός αυτό καθιστά τους "τυχαίους" κώδικες ένα καθαρά θεωρητικό εργαλείο.

3.2.2 Απόδειξη θεωρήματος Shannon: Αντίστροφο

Αντίστροφο: *Αν υπάρχει ακολουθία $(2^{nR}, n)$ κωδίκων με $\lambda^{(n)} \to 0$, όταν $n \to \infty$, τότε $R \leq C$.*

Παρατηρήστε ότι αν $\lambda^{(n)} \to 0$ όταν $n \to \infty$, τότε και $P_n^{(e)} \to 0$ όταν $n \to \infty$. Συνεπώς, αρκεί να αποδείξουμε ότι *αν υπάρχει ακολουθία $(2^{nR}, n)$ κωδίκων με $P_e^{(n)} \to 0$, όταν $n \to \infty$, τότε $R \leq C$.*

Αρχικά, θα αποδείξουμε ότι η χωρητικότητα (ανά χρήση καναλιού) δεν αυξάνεται αν χρησιμοποιήσουμε ένα διακριτό κανάλι χωρίς μνήμη πολλές φορές.

Λήμμα 1 *Έστω Y^n η έξοδος ενός διακριτού καναλιού χωρίς μνήμη με είσοδο X^n. Τότε:*

$$I(X^n; Y^n) \leq n\, C \tag{3.24}$$

για όλες τις συναρτήσεις μάζας πιθανότητας $p(x^n)$.

Απόδειξη:

$$\begin{aligned} I(X^n;Y^n) &= H(Y^n) - H(Y^n|X^n) \\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1,\ldots,Y_{i-1},X^n) \\ &\overset{(a)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &= \sum_{i=1}^{n} I(Y_i;X_i) \\ &\overset{(b)}{\leq} n\,C \end{aligned} \tag{3.25}$$

όπου στο σημείο (a) χρησιμοποιήσαμε την "αμνησία" του καναλιού και στο σημείο (b) τον ορισμό της χωρητικότητας. □

Ένα βασικό εργαλείο για την απόδειξη του αντίστροφου του θεωρήματος κωδικοποίησης είναι η ανισότητα του Fano.

Ανισότητα Fano

Έστω η τυχαία μεταβλητή W, η οποία είναι ομοιόμορφα κατανεμημένη στο σύνολο $\mathcal{W} := \{1,2,\ldots,2^{nR}\}$, και η τυχαία μεταβλητή $\hat{W}$, η οποία είναι πιθανοτικά συνδεδεμένη με την W. Για παράδειγμα,

$$W \longrightarrow X^n(W) \longrightarrow Y^n \longrightarrow \hat{W} = g(Y^n). \tag{3.26}$$

Ορίζουμε την πιθανότητα σφάλματος εκτίμησης της W, μέσω της $\hat{W}$, ως

$$P(\hat{W} \neq W) := \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \lambda_i = P_e^{(n)} \tag{3.27}$$

και τη δυαδική τυχαία μεταβλητή

$$E = \begin{cases} 1, & \text{αν } \hat{W} \neq W, \\ 0, & \text{αν } \hat{W} = W. \end{cases} \tag{3.28}$$

Χρησιμοποιώντας τον κανόνα της αλυσίδας, μπορούμε να γράψουμε

$$\begin{aligned} H(E,W|\hat{W}) &= H(W|\hat{W}) + \underbrace{H(E\,|\,W,\hat{W})}_{=0} \\ &= \underbrace{H(E\,|\,\hat{W})}_{\leq H(E)=H(P_e^{(n)})\leq 1} + H(W\,|\,E,\hat{W}). \end{aligned} \tag{3.29}$$

Επιπλέον

$$\begin{aligned} H(W|E,\hat{W}) &= P(E=0)\; H(W|\hat{W},E=0) + P(E=1)\; H(W|\hat{W},E=1) \\ &\leq P_e^{(n)} \log(|\mathcal{W}|-1) \leq P_e^{(n)}\, nR. \end{aligned} \tag{3.30}$$

Από τις (3.29) και (3.30), λαμβάνουμε μία έκφραση της ανισότητας Fano

$$H(W|\hat{W}) \leq 1 + P_e^{(n)}\, nR. \tag{3.31}$$

Απόδειξη αντίστροφου θεωρήματος Shannon

Για δεδομένο κώδικα και δεδομένη συνάρτηση αποκωδικοποίησης $\hat{W} = g(Y^n)$, έχουμε

$$W \to X^n(W) \to Y^n \to \hat{W}.$$

Συνεπώς

$$\begin{aligned} nR = H(W) &= H(W|\hat{W}) + I(W;\hat{W}) \\ &\overset{(a)}{\leq} 1 + P_e^{(n)}\, nR + I(W;\hat{W}) \\ &\overset{(b)}{\leq} 1 + P_e^{(n)}\, nR + I(X^n;Y^n) \\ &\overset{(c)}{\leq} 1 + P_e^{(n)}\, nR + nC \end{aligned} \tag{3.32}$$

όπου στο σημείο (a) χρησιμοποιήσαμε την ανισότητα Fano, στο (b) την ανισότητα επεξεργασίας δεδομένων, και στο (c) τη σχέση (3.24). Διαιρώντας και τα δύο μέλη της (3.32) με το n, λαμβάνουμε

$$R \leq P_e^{(n)} R + \frac{1}{n} + C. \tag{3.33}$$

Παίρνοντας όριο για $n \to \infty$, λαμβάνουμε[4] $R \leq C$.

[4]Θα μπορούσαμε να είχαμε ξεκινήσει στην (3.32) από οποιαδήποτε κατανομή για το W. Τότε, χρησιμοποιώντας τη σχέση $P_e^{(n)} \leq \lambda^{(n)}$ θα είχαμε καταλήξει στη σχέση

$$\frac{1}{n} H(W) \leq 1 + \lambda^{(n)} R + C.$$

Παίρνοντας $\sup_{p_W(w)}$ και $\lim_{n\to\infty}$ θα καταλήγαμε πάλι στο $R \leq C$.

Μπορούμε να γράψουμε την (3.33) ως:

$$P_e^{(n)} \geq 1 - \frac{C}{R} - \frac{1}{nR}.$$

Η σχέση αυτή υποδεικνύει ότι, αν $R > C$, τότε η πιθανότητα σφάλματος φράσσεται μακριά από το 0 για αρκετά μεγάλο n. Άρα, δεν μπορούμε να επιτύχουμε αυθαίρετα μικρή πιθανότητα σφάλματος για ρυθμούς μεγαλύτερους της χωρητικότητας. Το παραπάνω αποτέλεσμα καλείται ασθενές αντίστροφο (weak converse). Μπορεί να αποδειχθεί το ισχυρό αντίστροφο, το οποίο δηλώνει ότι, για ρυθμούς μεγαλύτερους της χωρητικότητας, η πιθανότητα σφάλματος τείνει εκθετικά στο 1.

Ισότητα στο αντίστροφο: Για κώδικες μηδενικού σφάλματος ($P_e^{(n)} = 0$):

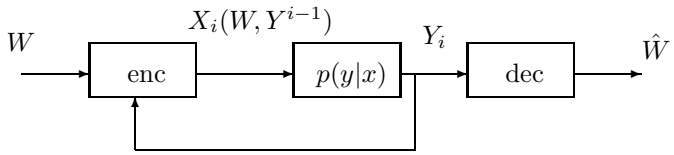
$$\begin{aligned}
nR &= H(W) \\
&= H(W|\hat{W}) + I(W;\hat{W}) \\
&= I(W;\hat{W}) \\
&\overset{(a)}{\leq} I(X^n(W);Y^n) \\
&= H(Y^n) - H(Y^n|X^n) \\
&= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \quad \text{(κανάλι DMC)} \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\
&= \sum_{i=1}^{n} I(X_i;Y_i) \\
&\overset{(c)}{\leq} n\,C.
\end{aligned} \tag{3.34}$$

Για να ισχύει ισότητα στην (a) πρέπει όλες οι κωδικές λέξεις είναι διαφορετικές, στην (b) πρέπει τα Y_i να είναι ανεξάρτητα, και στην (c) πρέπει η $p(x)$ να επιτυγχάνει τη χωρητικότητα.

3.3 Χωρητικότητα με ανάδραση (Feedback capacity)

Σε αυτό το εδάφιο μελετούμε το ερώτημα: *Μπορούμε να επιτύχουμε μεγαλύτερη χωρητικότητα χρησιμοποιώντας ανάδραση;* Η απάντηση προκαλεί έκπληξη και είναι Όχι!

Ένας $(2^{nR}, n)$ κώδικας με ανάδραση (feedback code) αποτελείται από



Σχήμα 3.4: Διάγραμμα επικοινωνιακού συστήματος με ανάδραση

1. την ακολουθία συναρτήσεων κωδικοποίησης $x_i(W, Y^{i-1})$, όπου κάθε x_i είναι συνάρτηση του μυνήματος W και όλων των προηγούμενων εξόδων $Y_1, \ldots, Y_{i-1}$,

2. τη συνάρτηση αποκωδικοποίησης $g : \mathcal{Y}^n \to \{1, 2, \ldots, 2^{nR}\}$.

Τότε:

$$P_e^{(n)} = \Pr\{g(Y^n) \neq W\}.$$

όταν W ομοιόμορφα κατανεμημένο στο σύνολο $\{1, 2, \ldots, 2^{nR}\}$. Η έννοια του επιτεύξιμου ρυθμού είναι ίδια με αυτή που ήδη αναφέρθηκε στο Θεώρημα Κωδικοποίησης χωρίς ανάδραση.

Ορισμός 13 *Η χωρητικότητα με χρήση ανάδρασης, C_{FB}, ενός διακριτού καναλιού χωρίς μνήμη είναι το ελάχιστο άνω φράγμα* (sup) *των επιτεύξιμων κωδίκων με ανάδραση.* □

Θεώρημα 3

$$C_{FB} = C = \max_{p(x)} I(X; Y). \tag{3.35}$$

Απόδειξη: Αφού κάθε κώδικας χωρίς ανάδραση είναι ειδική περίπτωση ενός κώδικα με ανάδραση, $C_{\mathrm{FB}} \geq C$. Για να αποδείξουμε την αντίστροφη ανισότητα, θα αποδείξουμε μία σχέση αντίστοιχη με το (ασθενές) αντίστροφο του θεωρήματος κωδικοποίησης. Δηλαδή ότι αν $P_e^{(n)} \to 0$, τότε $R \leq C = \max_{p(x)} I(X; Y)$. Δεν μπορούμε να χρησιμοποιήσουμε την ίδια απόδειξη με αυτή του αντίστροφου θεωρήματος κωδικοποίησης και ειδικά τη σχέση

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

διότι το X_i εξαρτάται από παρελθόντα σύμβολα εξόδου και άρα το Y_i συνδέεται πιθανοτικά με X_j, για $j \geq i$. Πάντως, μία μικρή αλλαγή αρκεί. Θα χρησιμοποιήσουμε τον δείκτη W αντί για το X^n και θα ακολουθήσουμε μία αντίστοιχη αλυσίδα ανισοτήτων.

Έστω W ομοιόμορφα κατανεμημένο στο $\{1, 2, \ldots, 2^{nR}\}$. Τότε:

$$\begin{aligned} nR = H(W) &= H(W|\hat{W}) + I(W; \hat{W}) \\ &\overset{(a)}{\leq} 1 + P_e^{(n)}\, nR + I(W; \hat{W}) \\ &\overset{(b)}{\leq} 1 + P_e^{(n)}\, nR + I(W; Y^n) \end{aligned} \tag{3.36}$$

όπου στο σημείο (a) χρησιμοποιήσαμε την ανισότητα Fano και στο σημείο (b) την ανισότητα επεξεργασίας δεδομένων. Επιπλέον

$$\begin{aligned} I(W;Y^n) &= H(Y^n) - H(Y^n|W) \\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1,\ldots,Y_{i-1},W) \\ &\overset{(a)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1,Y_2,\ldots,Y_{i-1},W,X_i) \\ &\overset{(b)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &= \sum_{i=1}^{n} I(X_i;Y_i) \\ &\leq n\,C \end{aligned} \tag{3.37}$$

όπου στο σημείο (a) χρησιμοποιούμε το ότι $X_i = X_i(W, Y^{i-1})$, και στο σημείο (b) το γεγονός ότι, υπό συνθήκη $X_i(W, Y^{i-1})$, το Y_i είναι ανεξάρτητο του W και των $Y_1, \ldots, Y_{i-1}$.

Από τις (3.36) και (3.37) λαμβάνουμε ότι

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

Διαιρώντας με n και παίρνοντας όριο για $n \to \infty$, λαμβάνουμε $R \leq C$. □

Κεφάλαιο 4

Θεωρία Ρυθμού Παραμόρφωσης

4.1 Εισαγωγή - Ορισμοί

Σε αυτό το κεφάλαιο, θα αποδείξουμε το δεύτερο σημαντικό θεώρημα του Shannon, το Θεώρημα Ρυθμού Παραμόρφωσης.

Έστω πηγή πληροφορίας η οποία, κάθε χρονική στιγμή, παράγει ανεξάρτητα και όμοια κατανεμημένα τυχαία σύμβολα X, από το πεπερασμένο αλφάβητο πηγής $\mathcal{X}$, με συνάρτηση μάζας πιθανότητας $p_X(x) = P(X = x)$, για $x \in \mathcal{X}$. Η πηγή αυτή καλείται *διακριτή πηγή χωρίς μνήμη* (discrete memoryless source). Ακολουθίες συμβόλων που παράγονται από την πηγή συμβολίζονται ως $X_1, X_2, \ldots$.

Έστω ότι επιθυμούμε να μεταδώσουμε την έξοδο της πηγής σε έναν προορισμό, μέσω ενός καναλιού, και έστω ότι το σύμβολο $x \in \mathcal{X}$ θα αναπαραχθεί στον προορισμό ως $\hat{x} \in \hat{\mathcal{X}}$, όπου $\hat{\mathcal{X}}$ είναι το αλφάβητο προορισμού (σε πολλές περιπτώσεις, $\mathcal{X} = \hat{\mathcal{X}}$).

Ορισμός 14 *Για κάθε ζεύγος $(x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}$ ορίζουμε τη συνάρτηση (μέτρο) παραμόρφωσης (distortion function (measure)), $d(x, \hat{x})$, με*

$$d : \mathcal{X} \times \hat{\mathcal{X}} \longrightarrow R^+, \tag{4.1}$$

η οποία μετράει το **σφάλμα** *(error) ή* **παραμόρφωση** *(distortion) που προκύπτει από την αναπαράσταση του x από το $\hat{x}$.* □

Παράδειγμα: Η **παραμόρφωση Hamming** (Hamming distortion, Hamming distance) ορίζεται ως εξής:

$$d(x, \hat{x}) = \begin{cases} 0, & \text{αν } x = \hat{x}, \\ 1, & \text{αν } x \neq \hat{x} \end{cases} . \tag{4.2}$$

Η μέση παραμόρφωση Hamming ισούται με την πιθανότητα σφάλματος, $P(X \neq \hat{X})$. Πράγματι

$$\mathcal{E}d(X,\hat{X}) = \sum_{(x,\hat{x})\in\mathcal{X}\times\hat{\mathcal{X}}} p(x,\hat{x})d(x,\hat{x}) = \sum_{(x,\hat{x}):x\neq\hat{x}} p(x,\hat{x}) = P(X \neq \hat{X}). \tag{4.3}$$

$\diamond$

Ορισμός 15 *Η συνάρτηση παραμόρφωσης* $d : \mathcal{X} \times \hat{\mathcal{X}} \to R^+$ *καλείται* **φραγμένη** *(bounded) αν*

$$d_{\max} := \max_{x\in\mathcal{X},\hat{x}\in\hat{\mathcal{X}}} d(x,\hat{x}) < \infty. \tag{4.4}$$

□

Ορισμός 16 *Για κάθε ζεύγος ακολουθιών* $(x^n, \hat{x}^n) \in \mathcal{X}^n \times \hat{\mathcal{X}}^n$ *ορίζουμε τη συνάρτηση παραμόρφωσης ακολουθίας*

$$d(x^n,\hat{x}^n) := \frac{1}{n}\sum_{i=1}^{n} d(x_i,\hat{x}_i). \tag{4.5}$$

□

Δηλαδή, η παραμόρφωση ακολουθίας ισούται με τον αριθμητικό μέσο των παραμορφώσεων των συμβόλων της ακολουθίας.

Ορισμός 17 *Ένας* $\left(2^{nR}, n\right)$ *κώδικας ρυθμού παραμόρφωσης αποτελείται από*

1. *μία συνάρτηση κωδικοποίησης*

$$f_n : \mathcal{X}^n \longrightarrow \left\{1, 2, \ldots, 2^{nR}\right\}, \tag{4.6}$$

2. *μία συνάρτηση αποκωδικοποίησης*

$$g_n : \left\{1, 2, \ldots, 2^{nR}\right\} \longrightarrow \hat{\mathcal{X}}^n. \tag{4.7}$$

Η παραμόρφωση που εισάγει ο κώδικας (f_n, g_n) *ορίζεται ως*

$$\begin{aligned} D(f_n,g_n) &:= \mathcal{E}_{X^n} d(X^n, g_n(f_n(X^n))) \\ &= \sum_{x^n\in\mathcal{X}^n} p(x^n)d(x^n, g_n(f_n(x^n))). \end{aligned} \tag{4.8}$$

□

Δηλαδή, όταν αναφερόμαστε στην παραμόρφωση που εισάγει ένας κώδικας, αναφερόμαστε στη μέση παραμόρφωση, πάνω σε όλες τις δυνατές εισόδους.

Ορισμός 18 *Ως κωδικό βιβλίο,* $\mathtt{C}$*, ενός συγκεκριμένου κώδικα ρυθμού παραμόρφωσης,* (f_n, g_n)*, ορίζουμε το σύνολο των κωδικών λέξεων, δηλαδή,* $\mathtt{C} := \left[g_n(1), \ldots, g_n(2^{nR})\right]$*.* □

Πολλές φορές, για απλούστευση συμβολισμού, το κωδικό βιβλίο ενός συγκεκριμένου κώδικα και ο ίδιος ο κώδικας θα συμβολίζονται με $\mathtt{C} := \left[\hat{x}^n(1), \ldots, \hat{x}^n(2^{nR})\right]$ και η αντίστοιχη παραμόρφωση θα συμβολίζεται με $D(\mathtt{C})$. Ένας τυχαίος κώδικας και η αντίστοιχη παραμόρφωση θα συμβολίζονται με $\mathcal{C} := \left[\hat{X}^n(1), \ldots, \hat{X}^n(2^{nR})\right]$ και $D(\mathcal{C})$, αντίστοιχα.

Ορισμός 19 *Το ζεύγος ρυθμού παραμόρφωσης* (R, D) *καλείται* **επιτεύξιμο** *(achievable) αν υπάρχει ακολουθία* $(2^{nR}, n)$ *κωδίκων ρυθμού παραμόρφωσης* (f_n, g_n) *με*

$$\lim_{n\to\infty} \mathcal{E}_{X^n} d(X^n, g_n(f_n(X^n)) \leq D. \tag{4.9}$$

□

Ορισμός 20 *Η κλειστή θήκη (closure) των επιτεύξιμων ζευγών ρυθμού παραμόρφωσης για μία πηγή καλείται* **περιοχή ρυθμού παραμόρφωσης** *(rate distortion region) της πηγής.* □

Ορισμός 21 *Έστω* $D \geq 0$*. Το μέγιστο κάτω φράγμα (infimum) των ρυθμών* R *για τους οποίους το ζεύγος* (R, D) *ανήκει στην περιοχή ρυθμού παραμόρφωσης μίας πηγής συμβολίζεται με* $R(D)$ *και καλείται* **συνάρτηση ρυθμού παραμόρφωσης** *(rate distortion function) της πηγής.* □

Ορισμός 22 *Έστω* $R \geq 0$*. Το μέγιστο κάτω φράγμα των παραμορφώσεων* D *για τις οποίες το ζεύγος* (R, D) *ανήκει στην περιοχή ρυθμού παραμόρφωσης μίας πηγής συμβολίζεται με* $D(R)$ *και καλείται* **συνάρτηση παραμόρφωσης ρυθμού** *(distortion rate function) της πηγής.* □

Ορισμός 23 *Έστω διακριτή πηγή χωρίς μνήμη, με αλφάβητο πηγής* $\mathcal{X}$*, συνάρτηση μάζας πιθανότητας* $p(x)$*, αλφάβητο προορισμού* $\hat{\mathcal{X}}$*, και μέτρο παραμόρφωσης* $d(x, \hat{x})$*. Η συνάρτηση* **ρυθμού παραμόρφωσης πληροφορίας** *(information rate distortion)* $R^{(I)}(D)$ *της πηγής ορίζεται ως εξής:*

$$R^{(I)}(D) := \min_{p(\hat{x}|x):\sum_{x,\hat{x}} p(x)p(\hat{x}|x)d(x,\hat{x})\leq D} I(X; \hat{X}). \tag{4.10}$$

□

Παρατήρηση: Προφανώς, $\sum_{x,\hat{x}} p(x)p(\hat{x}|x)d(x,\hat{x}) = \mathcal{E}d(X,\hat{X})$, συνεπώς, μία λίγο διαφορετικά διατυπωμένη έκφραση για το ρυθμό παραμόρφωσης πληροφορίας είναι η εξής:

$$R^{(I)}(D) := \min_{p(\hat{x}|x):\mathcal{E}d(X,\hat{X})\leq D} I(X;\hat{X}).$$

Θεώρημα 4 *Η συνάρτηση ρυθμού παραμόρφωσης πληροφορίας, $R^{(I)}(D)$, για δυαδική πηγή X, με κατανομή Bernoulli(p) και μέτρο παραμόρφωσης Hamming, ισούται με*

$$R^{(I)}(D) = \begin{cases} H(p) - H(D), & 0 \leq D \leq \min\{p, 1-p\}, \\ 0, & D > \min\{p, 1-p\}. \end{cases} \tag{4.11}$$

Απόδειξη: Χωρίς βλάβη της γενικότητας, έστω $p \leq \frac{1}{2}$. Έστω ζεύγος τυχαίων μεταβλητών $(X,\hat{X})$ με από-κοινού συνάρτηση μάζας πιθανότητας $p_{X,\hat{X}}(x,\hat{x}) = p_X(x)p_{\hat{X}|X}(\hat{x}|x)$, τέτοια ώστε $p_X(1) = p$, $p_X(0) = 1-p$, και $\mathcal{E}d(X,\hat{X}) = P(X \neq \hat{X}) \leq D$, με $0 \leq D \leq p$. Αφού η p_X είναι δεδομένη, ο περιορισμός για την παραμόρφωση αφορά στην $p_{\hat{X}|X}$, η οποία θα πρέπει να είναι τέτοια ώστε $\sum_{x,\hat{x}} p(x)p(\hat{x}|x)d(x,\hat{x}) \leq D$. Στη συνέχεια, αντί να υπολογίσουμε το $R^{(I)}(D)$ λύνοντας το πρόβλημα ελαχιστοποίησης (4.10), αρχικά, θα υπολογίσουμε ένα κάτω φράγμα για το $R^{(I)}(D)$ και, κατόπιν, θα αποδείξουμε ότι είναι επιτεύξιμο.

Έστω ότι το σύμβολο $\oplus$ συμβολίζει την πράξη "άθροισμα modulo 2," και έστω $Y = X \oplus \hat{X}$. Η Y είναι δυαδική τυχαία μεταβλητή με $P(Y=1) = P(X \neq \hat{X}) \leq D$. Υπενθυμίζουμε ότι, για $0 \leq x \leq 1$, $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ και ότι η $H(x)$ είναι αύξουσα στο διάστημα $0 \leq x \leq \frac{1}{2}$. Συνεπώς, $H(Y) \leq H(D)$.

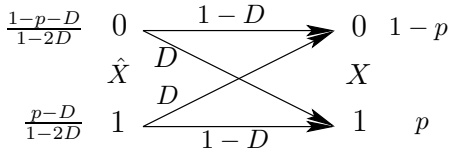
Για το ζεύγος $(X,\hat{X})$, έχουμε

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= H(p) - H(X \oplus \hat{X}|\hat{X}) \\ &\overset{(a)}{\geq} H(p) - H(X \oplus \hat{X}) \\ &\overset{(b)}{\geq} H(p) - H(D) \end{aligned} \tag{4.12}$$

όπου στο σημείο (a) χρησιμοποιήσαμε το ότι η συνθήκη δεν αυξάνει την εντροπία και στο σημείο (b) χρησιμοποιήσαμε το ότι $H(Y) \leq H(D)$.

Ελαχιστοποιώντας πάνω σε όλες τις $p(\hat{x}|x)$ οι οποίες ικανοποιούν τον περιορισμό παραμόρφωσης, λαμβάνουμε

$$\begin{aligned} &\min_{p(\hat{x}|x):\mathcal{E}d(X,\hat{X})\leq D} I(X;\hat{X}) \geq \min_{p(\hat{x}|x):\mathcal{E}d(X,\hat{X})\leq D} (H(p) - H(D)) \\ \implies \quad & R^{(I)}(D) \geq H(p) - H(D). \end{aligned} \tag{4.13}$$



Σχήμα 4.1: Κανάλι ελέγχου για το ρυθμό παραμόρφωσης δυαδικής πηγής χωρίς μνήμη, με συνάρτηση μάζας πιθανότητας Bernoulli(p).

Στη συνέχεια, βρίσκουμε μία από-κοινού συνάρτηση μάζας πιθανότητας $p(x, \hat{x})$, η οποία επιτυγχάνει το κάτω φράγμα της (4.13) και για την οποία $\mathcal{E}d(X, \hat{X}) \leq D$, αποδεικνύοντας το ένα σκέλος του Θεωρήματος. Έστω ζεύγος $(X, \hat{X})$ με $\hat{X}$ είσοδο στο ΔΣΚ του Σχήματος 4.1 και X έξοδο. Η πιθανότητα σφάλματος συμβόλου (crossover probability) ισούται με D, το οποίο ισοδυναμεί με $P(X \neq \hat{X}) = D$. Συνεπώς, παραμόρφωση $\leq D$ επιτυγχάνεται ανεξάρτητα από τη συνάρτηση μάζας πιθανότητας της εισόδου, $p_{\hat{X}}$.

Στη συνέχεια, επιλέγουμε την κατανομή της εισόδου $\hat{X}$ τέτοια ώστε η κατανομή της εξόδου X να είναι η Bernoulli(p). Έστω $P(\hat{X} = 1) = r$. Τότε

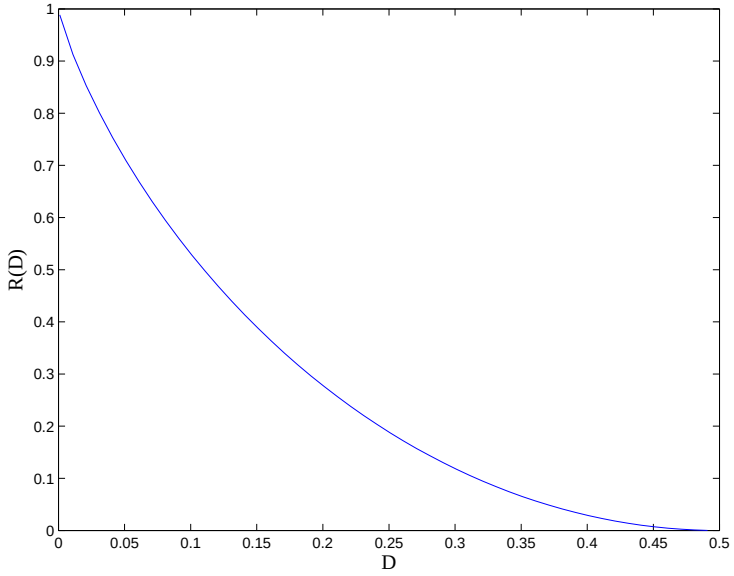
$$\begin{aligned} & p_X(1) = p_{\hat{X}}(0)p_{X|\hat{X}}(1|0) + p_{\hat{X}}(1)p_{X|\hat{X}}(1|1) \\ \Longrightarrow \quad & p = (1-r)D + r(1-D) \\ \Longrightarrow \quad & r = \frac{p-D}{1-2D}. \end{aligned} \tag{4.14}$$

Τέλος

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= H(p) - H(D). \end{aligned} \tag{4.15}$$

Συνεπώς, για τη συνάρτηση μάζας πιθανότητας $p_{X,\hat{X}} = p_{\hat{X}}p_{X|\hat{X}}$, έχουμε $X \sim$ Bernoulli(p), $\mathcal{E}d(X, \hat{X}) \leq D$, και $I(X;\hat{X}) = H(p) - H(d)$, αποδεικνύοντας την επιτευξιμότητα του κάτω φράγματος της (4.13).

Αν $D \geq p$, τότε μπορούμε να επιτύχουμε $R(D) = 0$ θέτοντας $\hat{X} = 0$ με πιθανότητα 1. Σε αυτή την περίπτωση, $I(X;\hat{X}) = 0$ και $D = p$. Όμοια, αν $D \geq 1 - p$, τότε θέτοντας $\hat{X} = 1$ με πιθανότητα 1, επιτυγχάνουμε $R(D) = 0$. Η $R(D)$ σχεδιάζεται στο Σχήμα 4.2. □



Σχήμα 4.2: Συνάρτηση ρυθμού παραμόρφωσης για $p = \frac{1}{2}$.

4.2 Τυπικό σύνολο ως προς την παραμόρφωση

Ορισμός 24 *Έστω $p(x, \hat{x})$ από-κοινού συνάρτηση μάζας πιθανότητας στο $\mathcal{X} \times \hat{\mathcal{X}}$, $d(x, \hat{x})$ μέτρο παραμόρφωσης στο $\mathcal{X} \times \hat{\mathcal{X}}$, και $\epsilon > 0$. Το ζεύγος ακολουθιών $(x^n, \hat{x}^n)$ καλείται ϵ-τυπικό ως προς την παραμόρφωση αν*

$$\begin{aligned} \left| -\frac{1}{n} \log p(x^n) - H(X) \right| &< \epsilon, \\ \left| -\frac{1}{n} \log p(\hat{x}^n) - H(\hat{X}) \right| &< \epsilon, \\ \left| -\frac{1}{n} \log p(x^n, \hat{x}^n) - H(X, \hat{X}) \right| &< \epsilon, \\ \left| d(x^n, \hat{x}^n) - \mathcal{E}d(X, \hat{X}) \right| &< \epsilon. \end{aligned} \tag{4.16}$$

Το σύνολο των ϵ-τυπικών ακολουθιών ως προς την παραμόρφωση καλείται **ϵ-τυπικό σύνολο ως προς την παραμόρφωση** *(ϵ-distortion typical set) και συμβολίζεται με $A_{d,\epsilon}^{(n)}$.*[1] ◇

Λήμμα 2 *Έστω $(X_i, \hat{X}_i)$ επιλέγονται ανεξάρτητα, όμοια κατανεμημένα, από την $p(x, \hat{x})$. Τότε, $P\left(A_{d,\epsilon}^{(n)}\right) \to 1$ όταν $n \to \infty$.*

Απόδειξη: Η απόδειξη αφήνεται σαν άσκηση. △

[1] Για απλούστευση συμβολισμού, συνήθως θα παραλείπουμε το ϵ, το οποίο όμως πάντα θα υπονοείται.

Λήμμα 3 *Αν* $(x^n, \hat{x}^n) \in A_{d,\epsilon}^{(n)}$, *τότε*

$$p(\hat{x}^n) \geq p(\hat{x}^n|x^n)\, 2^{-n(I(X;\hat{X})+3\epsilon)}. \tag{4.17}$$

Απόδειξη: Αν $(x^n, \hat{x}^n) \in A_{d,\epsilon}^{(n)}$, τότε μπορούμε να φράξουμε από πάνω και από κάτω τις πιθανότητες $p(x^n)$, $p(\hat{x}^n)$ και $p(x^n, \hat{x}^n)$. Συνεπώς

$$\begin{aligned} p(\hat{x}^n|x^n) &= \frac{p(x^n, \hat{x}^n)}{p(x^n)} \\ &= p(\hat{x}^n)\, \frac{p(x^n, \hat{x}^n)}{p(\hat{x}^n)\, p(x^n)} \\ &\leq p(\hat{x}^n)\, \frac{2^{-n(H(X,\hat{X})-\epsilon)}}{2^{-n(H(X)+\epsilon)}\, 2^{-n(H(\hat{X})+\epsilon)}} \\ &= p(\hat{x}^n)\, 2^{n(I(X;\hat{X})+3\epsilon)}. \end{aligned} \tag{4.18}$$

Το λήμμα αποδείχθηκε. △

Στη συνέχεια, θα φανεί χρήσιμο το παρακάτω αποτέλεσμα.

Λήμμα 4 *Αν* $0 \leq x, y \leq 1$ *και* $n > 0$, *τότε*

$$(1 - xy)^n \leq 1 - x + e^{-ny}. \tag{4.19}$$

Απόδειξη: Δείτε Cover&Thomas, σελ. 320. △

4.3 Θεώρημα Ρυθμού Παραμόρφωσης

Θεώρημα 5 *Η συνάρτηση ρυθμού παραμόρφωσης,* $R(D)$, *για μία πηγή η οποία παράγει ανεξάρτητα, όμοια κατανεμημένα, σύμβολα* $X \in \mathcal{X}$, *με συνάρτηση μάζας πιθανότητας* $p(x)$, *αλφάβητο προορισμού* $\hat{\mathcal{X}}$, *και φραγμένη συνάρτηση παραμόρφωσης* $d(x, \hat{x})$, *ισούται με τη συνάρτηση ρυθμού παραμόρφωσης πληροφορίας για την πηγή,* $R^{(I)}(D)$. *Δηλαδή,*

$$R(D) = R^{(I)}(D). \tag{4.20}$$

□

Στη συνέχεια, θα αποδείξουμε το σημαντικό αυτό θεώρημα. Αρχικά, θα αποδείξουμε την επιτευξιμότητα του ζεύγους (R, D), για κάθε $R > R^{(I)}(D)$ και, κατόπιν, θα αποδείξουμε ότι, αν $R < R^{(I)}(D)$, τότε το ζεύγος (R, D) δεν είναι επιτεύξιμο.

4.3.1 Απόδειξη επιτευξιμότητας

1. *Δημιουργία κωδικού βιβλίου.* Έστω υπό-συνθήκη συνάρτηση μάζας πιθανότητας $p(\hat{x}|x)$, τέτοια ώστε
$$\sum_{(x,\hat{x})\in\mathcal{X}\times\hat{\mathcal{X}}} p(x)p(\hat{x}|x)d(x,\hat{x}) \leq D. \tag{4.21}$$
Συνεπώς, η συγκεκριμένη υπό-συνθήκη μάζα πιθανότητας, $p(\hat{x}|x)$, ανήκει στο σύνολο πάνω στο οποίο ελαχιστοποιούμε την αμοιβαία πληροφορία $I(X;\hat{X})$ στην (4.10).

   Δημιουργούμε τυχαίο κώδικα, $\mathcal{C}$, με 2^{nR} κωδικές λέξεις, $\hat{X}^n(i)$, για $i = 1,\ldots,2^{nR}$, μήκους n η καθεμία, επιλέγοντας τα στοιχεία κάθε κωδικής λέξης τυχαία, ανεξάρτητα, όμοια κατανεμημένα, με συνάρτηση μάζας πιθανότητας $p(\hat{x}) = \sum_x p(x)p(\hat{x}|x)$. Έστω συγκεκριμένος κώδικας $\mathtt{C} = [\hat{x}^n(1),\ldots,\hat{x}^n(2^{nR})]$. Η πιθανότητα του $\mathtt{C}$ ισούται με
$$P(\mathtt{C}) = \prod_{i=1}^{2^{nR}} p(\hat{x}^n(i)) \tag{4.22}$$
   όπου αν $\hat{x}^n(i) = [\hat{x}^n(i,1)\ \cdots\ \hat{x}^n(i,n)]$, για $i = 1,\ldots,2^{nR}$, τότε $p(\hat{x}^n(i)) = \prod_{j=1}^n p(\hat{x}(i,j))$.

2. *Κωδικοποίηση.* Αντιστοιχίζουμε την είσοδο X^n στο δείκτη W αν $(X^n,\hat{X}^n(W)) \in A_{d,\epsilon}^{(n)}$. Αν υπάρχουν παραπάνω από ένα τέτοια W, τότε το X^n αντιστοιχίζεται στο μικρότερο από αυτά. Αν δεν υπάρχει κανένα τέτοιο W, τότε θέτουμε αυθαίρετα μία τιμή, έστω $W = 1$. Παρατηρούμε ότι nR bits αρκούν για να περιγράψουμε τη X^n.

3. *Αποκωδικοποίηση.* Η ακολουθία που αναπαράγουμε είναι η $\hat{X}^n(W)$.

Θα αποδείξουμε ότι, αν $R > I(X;\hat{X})$, τότε υπάρχει κώδικας ρυθμού παραμόρφωσης με ρυθμό R και ασυμπτωτική, ως προς το n, παραμόρφωση $\leq D$. Αν ως $p(\hat{x}|x)$ επιλέξουμε την υπό-συνθήκη συνάρτηση μάζας πιθανότητας η οποία ελαχιστοποιεί την $I(X;\hat{X})$ στην (4.10), τότε, αν $R > R^{(I)}(D)$, υπάρχει κώδικας ρυθμού παραμόρφωσης με ρυθμό R και ασυμπτωτική, ως προς το n, παραμόρφωση $\leq D$, αποδεικνύοντας την επιτευξιμότητα του ζεύγους $(R^{(I)}(D), D)$.

Η απόδειξη θα γίνει σε βήματα.

1. Έστω *δεδομένος* κώδικας $\mathtt{C} = \left[\hat{x}^n(1),\ldots,\hat{x}^n(2^{nR})\right]$, $\epsilon > 0$, και *τυχαία* είσοδος X^n. Η παραμόρφωση που εισάγει ο κώδικας $\mathtt{C}$ ισούται με
$$\begin{aligned} d(\mathtt{C}) &= \mathcal{E}_{X^n} d(X^n,\hat{X}^n) \\ &= \sum_{x^n\in\mathcal{X}^n} p(x^n)\, d(x^n, g_n(f_n(x^n))). \end{aligned} \tag{4.23}$$

Έστω $J(\mathtt{C})$ το σύνολο των εισόδων $x^n \in \mathcal{X}^n$ για τις οποίες υπάρχει κωδική λέξη του κώδικα $\mathtt{C}$, έστω $\hat{x}^n(i)$, τέτοια ώστε $(x^n, \hat{x}^n(i)) \in A_{d,\epsilon}^{(n)}$ (προφανώς, $J(\mathtt{C}) \subseteq \mathcal{X}^n$). Τότε

$$d(\mathtt{C}) = \sum_{x^n \in J(\mathtt{C})} p(x^n)\, d(x^n, g_n(f_n(x^n))) + \sum_{x^n \notin J(\mathtt{C})} p(x^n)\, d(x^n, g_n(f_n(x^n))). \tag{4.24}$$

Το πρώτο άθροισμα είναι $\leq D + \epsilon$ (γιατί;). Αν $d_{\max}$ είναι η μέγιστη παραμόρφωση του κώδικα, τότε

$$d(\mathtt{C}) \leq D + \epsilon + d_{\max} \sum_{x^n \notin J(\mathtt{C})} p(x^n). \tag{4.25}$$

Το άθροισμα στην παραπάνω σχέση ισούται με την πιθανότητα η είσοδος να μην μπορεί να αναπαρασταθεί καλά από τον κώδικα $\mathtt{C}$ (για ποιους λόγους;).

Ορίζουμε

$$\phi(x^n, \hat{x}^n) = \begin{cases} 1, & \text{αν } (x^n, \hat{x}^n) \in A_{d,\epsilon}^{(n)} \\ 0, & \text{αν } (x^n, \hat{x}^n) \notin A_{d,\epsilon}^{(n)} \end{cases}, \tag{4.26}$$

$$\phi^c(x^n, \hat{x}^n) = 1 - \phi(x^n, \hat{x}^n). \tag{4.27}$$

Τότε (προσπαθήστε να καταλάβετε την επόμενη πολύ σημαντική σχέση),

$$\sum_{x^n \notin J(\mathtt{C})} p(x^n) \stackrel{!}{=} \sum_{x^n \in \mathcal{X}^n} p(x^n) \prod_{i=1}^{2^{nR}} \phi^c(x^n, \hat{x}^n(i)). \tag{4.28}$$

Αν ορίσουμε

$$K(\mathtt{C}) = \sum_{x^n \in \mathcal{X}^n} p(x^n) \prod_{i=1}^{2^{nR}} \phi^c(x^n, \hat{x}^n(i)), \tag{4.29}$$

τότε

$$d(\mathtt{C}) \leq D + \epsilon + K(\mathtt{C})\, d_{\max}. \tag{4.30}$$

Συνεπώς, αν μπορούσαμε να βρούμε ένα κώδικα $\mathtt{C}$ με $K(\mathtt{C}) \leq 2\epsilon$ για αρκετά μεγάλο n, τότε η παραμόρφωση $D(\mathtt{C})$ του κώδικα θα ήταν $\leq D + \epsilon(1 + 2d_{\max})$, για αρκετά μεγάλο n. Αφού το ϵ είναι αυθαίρετο, αυτό θα σήμαινε ότι κατασκευάσαμε κώδικα ρυθμού παραμόρφωσης με ρυθμό R και ασυμπτωτική, ως προς το n, παραμόρφωση $\leq D$. Όμως, η εύρεση ενός τέτοιου κώδικα δεν είναι εύκολη. Αντ΄ αυτού, θα χρησιμοποιήσουμε ένα επιχείρημα τυχαίας κωδικοποίησης (random coding argument) και θα αποδείξουμε ότι, για κάθε $\epsilon > 0$, $R > I(X; \hat{X}) + 3\epsilon$ και αρκετά μεγάλο n, υπάρχει $(2^{nR}, n)$ κώδικας ρυθμού παραμόρφωσης, έστω $\mathtt{C}^*$, με $K(\mathtt{C}^*) \leq 2\epsilon$.

2. Θεωρούμε την τυχαία ποσότητα $\mathcal{K} = K(\mathcal{C})$ και υπολογίζουμε τη μέση τιμή της πάνω σε όλους τους δυνατούς κώδικες. Συνεπώς,

$$\begin{aligned}
\mathcal{E}_{\mathcal{C}}(\mathcal{K}) &= \sum_{\mathtt{c}\in\left(\hat{\mathcal{X}}^n\right)^{2^{nR}}} P(\mathtt{c})K(\mathtt{c}) \\
&= \sum_{(\hat{x}^n(1),\ldots,\hat{x}^n(2^{nR}))\in\left(\hat{\mathcal{X}}^n\right)^{2^{nR}}} p(\hat{x}^n(1))\cdots p(\hat{x}^n(2^{nR})) \sum_{x^n\in\mathcal{X}^n} p(x^n)\prod_{i=1}^{2^{nR}}\phi^c(x^n,\hat{x}^n(i)) \\
&= \sum_{x^n\in\mathcal{X}^n} p(x^n) \sum_{(\hat{x}^n(1),\ldots,\hat{x}^n(2^{nR}))\in\left(\hat{\mathcal{X}}^n\right)^{2^{nR}}} p(\hat{x}^n(1))\cdots p(\hat{x}^n(2^{nR}))\prod_{i=1}^{2^{nR}}\phi^c(x^n,\hat{x}^n(i)) \\
&= \sum_{x^n\in\mathcal{X}^n} p(x^n) \sum_{(\hat{x}^n(1),\ldots,\hat{x}^n(2^{nR}))\in\left(\hat{\mathcal{X}}^n\right)^{2^{nR}}} \prod_{i=1}^{2^{nR}} p(\hat{x}^n(i))\phi^c(x^n,\hat{x}^n(i)) \\
&\overset{(!)}{=} \sum_{x^n\in\mathcal{X}^n} p(x^n)\left(\sum_{\hat{x}^n\in\hat{\mathcal{X}}^n} p(\hat{x}^n)\phi^c(x^n,\hat{x}^n)\right)^{2^{nR}}
\end{aligned} \tag{4.31}$$

όπου στο σημείο (!) χρησιμοποιήσαμε τη σχέση

$$\left(\sum_{x\in\mathcal{A}} f(x)\right)^M = \sum_{x_1\in\mathcal{A}}\cdots\sum_{x_M\in\mathcal{A}} f(x_1)\cdots f(x_M), \tag{4.32}$$

όπου $\mathcal{A}$ πεπερασμένο σύνολο.

Από τον ορισμό του ϕ^c, λαμβάνουμε

$$\begin{aligned}
\sum_{\hat{x}^n\in\hat{\mathcal{X}}^n} p(\hat{x}^n)\phi^c(x^n,\hat{x}^n) &= \sum_{\hat{x}^n\in\hat{\mathcal{X}}^n} p(\hat{x}^n)(1-\phi(x^n,\hat{x}^n)) \\
&= 1 - \sum_{\hat{x}^n\in\hat{\mathcal{X}}^n} p(\hat{x}^n)\phi(x^n,\hat{x}^n).
\end{aligned} \tag{4.33}$$

Συνεπώς

$$\mathcal{E}_{\mathcal{C}}(\mathcal{K}) = \sum_{x^n\in\mathcal{X}^n} p(x^n)\left(1-\sum_{\hat{x}^n\in\hat{\mathcal{X}}^n} p(\hat{x}^n)\phi(x^n,\hat{x}^n)\right)^{2^{nR}}. \tag{4.34}$$

Αν $(x^n,\hat{x}^n)\in A_{d,\epsilon}^{(n)}$, τότε, από το Λήμμα 3, έχουμε

$$p(\hat{x}^n) \geq p(\hat{x}^n|x^n)2^{-n(I(X;\hat{X})+3\epsilon)} \tag{4.35}$$

συνεπώς

$$\begin{aligned} 1 - \sum_{\hat{x}^n \in \hat{\mathcal{X}}^n} p(\hat{x}^n)\phi(x^n, \hat{x}^n) &\leq 1 - \sum_{\hat{x}^n \in \hat{\mathcal{X}}^n} p(\hat{x}^n|x^n) 2^{-n(I(X;\hat{X})+3\epsilon)} \phi(x^n, \hat{x}^n) \\ &= 1 - 2^{-n(I(X;\hat{X})+3\epsilon)} \sum_{\hat{x}^n \in \hat{\mathcal{X}}^n} p(\hat{x}^n|x^n)\phi(x^n, \hat{x}^n) \end{aligned} \tag{4.36}$$

και (γιατί;)

$$\left(1 - \sum_{\hat{x}^n \in \hat{\mathcal{X}}^n} p(\hat{x}^n)\phi(x^n, \hat{x}^n)\right)^{2^{nR}} \leq \left(1 - 2^{-n(I(X;\hat{X})+3\epsilon)} \sum_{\hat{x}^n \in \hat{\mathcal{X}}^n} p(\hat{x}^n|x^n)\phi(x^n, \hat{x}^n)\right)^{2^{nR}}. \tag{4.37}$$

Χρησιμοποιώντας το Λήμμα 4, λαμβάνουμε

$$\begin{aligned} &\left(1 - 2^{-n(I(X;\hat{X})+3\epsilon)} \sum_{\hat{x}^n \in \hat{\mathcal{X}}^n} p(\hat{x}^n|x^n)\phi(x^n, \hat{x}^n)\right)^{2^{nR}} \\ &\qquad \leq 1 - \sum_{\hat{x}^n \in \hat{\mathcal{X}}^n} p(\hat{x}^n|x^n)\phi(x^n, \hat{x}^n) + e^{-2^{-n(I(X;\hat{X})+3\epsilon)2^{nR}}}. \end{aligned} \tag{4.38}$$

Συνεπώς

$$\begin{aligned} \mathcal{E}_{\mathcal{C}}(\mathcal{K}) &\leq 1 - \sum_{(x^n, \hat{x}^n) \in \mathcal{X}^n \times \hat{\mathcal{X}}^n} p(x^n)p(\hat{x}^n|x^n)\phi(x^n, \hat{x}^n) + e^{-2^{-n(I(X;\hat{X})+3\epsilon)2^{nR}}} \\ &= \sum_{(x^n, \hat{x}^n) \in \mathcal{X}^n \times \hat{\mathcal{X}}^n} p(x^n, \hat{x}^n)\phi^c(x^n, \hat{x}^n) + e^{-2^{n(R-I(X;\hat{X})-3\epsilon)}}. \end{aligned} \tag{4.39}$$

Ο πρώτος όρος είναι η πιθανότητα του μη-τυπικού συνόλου, η οποία είναι $< \epsilon$ για αρκετά μεγάλο n. Ο δεύτερος όρος μπορεί να γίνει αυθαίρετα μικρός αν $R > I(X;\hat{X}) + 3\epsilon$ και n αρκετά μεγάλο. Συνεπώς, αν $R > I(X;\hat{X}) + 3\epsilon$ και n αρκετά μεγάλο, τότε

$$\mathcal{E}_{\mathcal{C}}(\mathcal{K}) \leq 2\epsilon. \tag{4.40}$$

Συνεπώς, για κάθε $\epsilon > 0$ και αρκετά μεγάλο n, υπάρχει κώδικας ρυθμού παραμόρφωσης, έστω $\mathtt{C}^*$, με ρυθμό $R > I(X;\hat{X}) + 3\epsilon$ και $K(\mathtt{C}^*) \leq 2\epsilon$.

3. Υπενθυμίζουμε τη σχέση (4.30), η οποία παρέχει ένα άνω φράγμα για την παραμόρφωση ενός κώδικα $\mathtt{C}$

$$d(\mathtt{C}) \leq D + \epsilon + d_{\max} K(\mathtt{C}). \tag{4.41}$$

Θέτοντας $\mathtt{C} = \mathtt{C}^*$, λαμβάνουμε

$$d(\mathtt{C}^*) \leq D + \epsilon(1 + 2d_{\max}). \tag{4.42}$$

Συνεπώς, για κάθε $\epsilon > 0$ και n αρκετά μεγάλο, υπάρχει κώδικας ρυθμού παραμόρφωσης με ρυθμό $R > I(X;\hat{X}) + 3\epsilon$ και παραμόρφωση $\leq D + \epsilon(1 + 2\,d_{\max})$. Άρα, το ζεύγος $(I(X;\hat{X}), D)$ είναι επιτεύξιμο.

4. Επιλέγοντας ως $p(\hat{x}|x)$ την υπό-συνθήκη μάζα πιθανότητας η οποία ελαχιστοποιεί την $I(X;\hat{X})$ στην (4.10) συμπεραίνουμε ότι το ζεύγος ρυθμού παραμόρφωσης $(R^{(I)}(D), D)$ είναι επιτεύξιμο.

4.3.2 Απόδειξη αντιστρόφου

Έστω διακριτή πηγή χωρίς μνήμη η οποία παράγει ανεξάρτητα, όμοια κατανεμημένα, σύμβολα $X \in \mathcal{X}$, με συνάρτηση μάζας πιθανότητας $p(x)$, αλφάβητο προορισμού $\hat{\mathcal{X}}$, και συνάρτηση παραμόρφωσης $d(x, \hat{x})$.

Στο προηγούμενο εδάφιο, αποδείξαμε ότι, για δεδομένη παραμόρφωση D, οποιοσδήποτε ρυθμός R, με $R > R^{(I)}(D)$, είναι επιτεύξιμος. Υπενθυμίζουμε ότι

$$R^{(I)}(D) = \min_{p(\hat{x}|x):\mathcal{E}d(X;\hat{X})\leq D} I(X;\hat{X}). \tag{4.43}$$

Στη συνέχεια, θα αποδείξουμε ότι δεν μπορούμε να επιτύχουμε παραμόρφωση $\leq D$ αν κωδικοποιήσουμε την έξοδο της πηγής X με κώδικα ρυθμού παραμόρφωσης με ρυθμό $R < R^{(I)}(D)$. Για την ακρίβεια, θα αποδείξουμε την ισοδύναμη (μέσω αντιθετοαντιστροφής) πρόταση: Αν υπάρχει $(2^{nR}, n)$ κώδικας ρυθμού παραμόρφωσης με παραμόρφωση $\leq D$, τότε θα πρέπει να ισχύει $R \geq R^{(I)}(D)$.

Το επόμενο αποτέλεσμα θα χρειαστεί για την απόδειξη του αντιστρόφου.

Λήμμα 5 *Η συνάρτηση ρυθμού παραμόρφωσης $R^{(I)}(D)$ είναι είναι μη-αύξουσα και κυρτή συνάρτηση του D.*

Απόδειξη: Η αύξηση της τιμής του D συνεπάγεται ότι το σύνολο $p(\hat{x}|x)$ πάνω στο οποίο ελαχιστοποιούμε στην (4.10) δεν μειώνεται. Συνεπώς, η ελάχιστη τιμή $R^{(I)}(D)$ δεν αυξάνεται.

Για να αποδείξουμε την κυρτότητα της $R^{(I)}(D)$, θεωρούμε τις παραμορφώσεις D_1 και D_2, οι οποίες έστω ότι αντιστοιχούν στους ρυθμούς $R_1 = R^{(I)}(D_1)$ και $R_2 = R^{(I)}(D_2)$. Έστω

ότι αυτά τα ζεύγη ρυθμού παραμόρφωσης επιτυγχάνονται από τις από-κοινού συναρτήσεις μάζας πιθανότητας $p_1(x,\hat{x}) = p(x)p_1(\hat{x}|x)$ και $p_2(x,\hat{x}) = p(x)p_2(\hat{x}|x)$, αντίστοιχα. Έστω $p_\lambda(x,\hat{x}) = \lambda p_1(x,\hat{x}) + (1-\lambda)p_2(x,\hat{x})$, για $0 \leq \lambda \leq 1$. Η παραμόρφωση είναι γραμμική συνάρτηση της από-κοινού συνάρτησης μάζας πιθανότητας. Συνεπώς, η παραμόρφωση για την $p_\lambda(x,\hat{x})$ ισούται με (να το αποδείξετε)

$$D_{p_\lambda} = \lambda D_1 + (1-\lambda)D_2. \tag{4.44}$$

Έχουμε αποδείξει ότι για δεδομένη συνάρτηση μάζας πιθανότητας, $p(x)$, η αμοιβαία πληροφορία $I(X;\hat{X})$ είναι κυρτή συνάρτηση της $p(\hat{x}|x)$. Συνεπώς,

$$I_{p_\lambda}(X;\hat{X}) \leq \lambda I_{p_1}(X;\hat{X}) + (1-\lambda)I_{p_2}(X;\hat{X}). \tag{4.45}$$

Από τον ορισμό (4.10), έχουμε ότι

$$\begin{aligned} R^{(I)}(D_{p_\lambda}) &\leq I_{p_\lambda}(X;\hat{X}) \\ &\leq \lambda I_{p_1}(X;\hat{X}) + (1-\lambda)I_{p_2}(X;\hat{X}_2) \\ &= \lambda R^{(I)}(D_1) + (1-\lambda)R^{(I)}(D_2) \end{aligned} \tag{4.46}$$

το οποίο αποδεικνύει την κυρτότητα της $R^{(I)}(D)$. □

Απόδειξη αντιστρόφου: Έστω $(2^{nR}, n)$ κώδικας ρυθμού παραμόρφωσης (f_n, g_n), ο οποίος αναπαριστά την είσοδο X^n με την κωδική λέξη $\hat{X}^n = g_n(f_n(X^n))$. Έστω ότι, για τον

συγκεκριμένο κώδικα, $\mathcal{E}d(X^n, \hat{X}^n) \leq D$. Τότε, ισχύει η ακόλουθη αλυσίδα ανισοτήτων

$$\begin{aligned}
nR &\overset{(a)}{\geq} H(f_n(X^n)) \\
&\overset{(b)}{=} H(f_n(X^n)) - H(f_n(X^n)|X^n) \\
&= I(X^n; f_n(X^n)) \\
&\overset{(c)}{\geq} I(X^n; \hat{X}^n) \\
&= H(X^n) - H(X^n|\hat{X}^n) \\
&= \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i|\hat{X}^n, X_1, \ldots, X_1) \\
&\overset{(d)}{\geq} \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i|\hat{X}_i) \\
&= \sum_{i=1}^{n} I(X_i; \hat{X}_i) \\
&\overset{(e)}{\geq} \sum_{i=1}^{n} R\left(\mathcal{E}d(X_i, \hat{X}_i)\right) \\
&= n\left(\frac{1}{n}\sum_{i=1}^{n} R\left(\mathcal{E}d(X_i, \hat{X}_i)\right)\right) \\
&\overset{(f)}{\geq} nR\left(\frac{1}{n}\sum_{i=1}^{n} \mathcal{E}d(X_i, \hat{X}_i)\right) \\
&= nR\left(\mathcal{E}d(X^n, \hat{X}^n)\right) \\
&= nR(D),
\end{aligned} \tag{4.47}$$

Ερωτήματα:

1. Να προσπαθήσετε να καταλάβετε τις ανισότητες

2. Ποια είναι η από-κοινού συνάρτηση μάζας πιθανότητας $p(x_i, \hat{x}_i)$ βάσει της οποίας υπολογίζουμε τις αμοιβαίες πληροφορίες $I(X_i; \hat{X}_i)$;