Θέματα Αλγορίθμων

Αλγόριθμοι και Εφαρμογές στον Πραγματικό Κόσμο

Μεταπτυχιακό Μάθημα

3η Εβδομάδα: Εντροπία, Τυχαιότητα & Πληροφορία

Σπύρος Κοντογιάννης

kontog@cse.uoi.gr

Τμήμα Μηχανικών Η/Υ & Πληροφορικής

Πανεπιστήμιο Ιωαννίνων

Τετάρτη, 8-15 Μαρτίου 2017

Σκελετός Ομιλίας

Τυχαιότητα & Συμπίεση

- Ποιο νόμισμα «παράγει» περισσότερη **τυχαιότητα**;
  - Δίκαιο Νόμισμα $[H : \frac{1}{2} \ , \ T : \frac{1}{2}]$;
  - Πειραγμένο Νόμισμα $[H : \frac{1}{4} \ , \ T : \frac{3}{4}]$;
  - Μεροληπτικό Νόμισμα $[H : 0 \ , \ T : 1]$;

Τυχαιότητα & Συμπίεση

- Ποιο νόμισμα «παράγει» περισσότερη **τυχαιότητα**;
  - Δίκαιο Νόμισμα $[H : \frac{1}{2} \ , \ T : \frac{1}{2}]$;
  - Πειραγμένο Νόμισμα $[H : \frac{1}{4} \ , \ T : \frac{3}{4}]$;
  - Μεροληπτικό Νόμισμα $[H : 0 \ , \ T : 1]$;

- Ποια n-ψήφια δυαδική συμβολοσειρά μπορεί να **συμπιεστεί** περισσότερο;
  - Τυχαία συμβολοσειρά με ακριβώς $n/2$ εμφανίσεις του «1»;
  - Τυχαία συμβολοσειρά με ακριβώς $n/10$ εμφανίσεις του «1»;
  - Τυχαία συμβολοσειρά με 1 εμφάνιση του «1»;

Τι είναι η Θεωρία της Πληροφορίας;

- Εμφανίστηκε το 1948.

Τι είναι η Θεωρία της Πληροφορίας;

- Εμφανίστηκε το 1948.
- Αφορά τη μαθηματική θεμελίωση της θεωρίας της επικοινωνίας.

Τι είναι η Θεωρία της Πληροφορίας;

- Εμφανίστηκε το 1948.
- Αφορά τη μαθηματική θεμελίωση της θεωρίας της επικοινωνίας.
- Ποσοτικοποιεί την έννοια της «πληροφορίας».

Τι είναι η Θεωρία της Πληροφορίας;

- Εμφανίστηκε το 1948.
- Αφορά τη μαθηματική θεμελίωση της θεωρίας της επικοινωνίας.
- Ποσοτικοποιεί την έννοια της «πληροφορίας».
- Τρεις βασικές προκλήσεις:
  - **Συμπίεση δίχως απώλειες:** Αφαίρεση «περιττής» πληροφορίας.
  - **Συμπίεση με απώλειες:** Αφαίρεση «περιττής» πληροφορίας, αλλά και του «θορύβου».
  - **Διόρθωση σφαλμάτων:** Προσθήκη «πλεονάζουσας» πληροφορίας για διάγνωση και αντιμετώπιση σφαλμάτων, πχ, λόγω «θορύβου» του μέσου μετάδοσης.

Τι είναι Πληροφορία;

- Είναι αυτό που απομένει αν αφαιρέσουμε όλα τα «περιττά δεδομένα».

Τι είναι Πληροφορία;

- Είναι αυτό που απομένει αν αφαιρέσουμε όλα τα «περιττά δεδομένα».
- ΠΑΡΑΔΕΙΓΜΑ: Η συμβολοσειρά

  00000000000000000000

  είναι «κενή πληροφορίας», ενώ η τυχαία συμβολοσειρά

  01101011110100100010

  περιέχει τη «μέγιστη δυνατή πληροφορία», που χρειάζεται 20 bits για να αναπαρασταθεί.

Τι είναι Πληροφορία;

- Είναι αυτό που απομένει αν αφαιρέσουμε όλα τα «περιττά δεδομένα».
- ΠΑΡΑΔΕΙΓΜΑ: Η συμβολοσειρά

  00000000000000000000

  είναι «κενή πληροφορίας», ενώ η τυχαία συμβολοσειρά

  01101011110100100010

  περιέχει τη «μέγιστη δυνατή πληροφορία», που χρειάζεται 20 bits για να αναπαρασταθεί.

$\therefore$ Το «σύνολο της πληροφορίας» που εμπεριέχεται σε ένα μήνυμα (δυαδική συμβολοσειρά) αφορά το «συνολο της τυχαιότητας» που αυτό το μήνυμα περιλαμβάνει.

Τι είναι Πληροφορία;

- Είναι αυτό που απομένει αν αφαιρέσουμε όλα τα «περιττά δεδομένα».
- ΠΑΡΑΔΕΙΓΜΑ: Η συμβολοσειρά

  00000000000000000000

  είναι «κενή πληροφορίας», ενώ η τυχαία συμβολοσειρά

  01101011110100100010

  περιέχει τη «μέγιστη δυνατή πληροφορία», που χρειάζεται 20 bits για να αναπαρασταθεί.

$\therefore$ Το «σύνολο της πληροφορίας» που εμπεριέχεται σε ένα μήνυμα (δυαδική συμβολοσειρά) αφορά το «συνολο της τυχαιότητας» που αυτό το μήνυμα περιλαμβάνει.

- Στη Θεωρία Πληροφορίας, αυτή η *ποσοτικοποίηση του «βαθμού τυχαιότητας»* καλείται **εντροπία**.

Εντροπία (I)

- Μια θεμελιώδης έννοια της Θεωρίας Πληροφορίας.
- $X =$ **διακριτή** τυχαία μεταβλητή, που παίρνει τιμές από ένα πεδίο $dom(X)$ σύμφωνα με συγκεκριμένη συνάρτηση πιθανότητας $P : dom(X) \mapsto \Delta(dom(X))$ τ.ώ:
  $\forall x \in dom(X),\ P(x) = \mathbb{P}[X = x]$.

Εντροπία (Ι)

- Μια θεμελιώδης έννοια της Θεωρίας Πληροφορίας.
- $X =$ **διακριτή** τυχαία μεταβλητή, που παίρνει τιμές από ένα πεδίο $dom(X)$ σύμφωνα με συγκεκριμένη συνάρτηση πιθανότητας $P : dom(X) \mapsto \Delta(dom(X))$ τ.ώ: $\forall x \in dom(X),\ P(x) = \mathbb{P}[X = x]$.

ΟΡΙΣΜΟΣ 9.1: Εντροπία

[Mitzenmacher-Upfal (2005), Chapter 9]

- Η **εντροπία σε bits** μιας διακριτής τυχαίας μεταβλητής X ορίζεται ως εξής:

$$H(X) = -\sum_{x \in dom(X)} \mathbb{P}[X = x] \log_2(\mathbb{P}[X = x]) = \mathbb{E}\left[\log_2\left(\frac{1}{P(X)}\right)\right]$$

Εντροπία (I)

- Μια θεμελιώδης έννοια της Θεωρίας Πληροφορίας.
- $X =$ **διακριτή** τυχαία μεταβλητή, που παίρνει τιμές από ένα πεδίο $dom(X)$ σύμφωνα με συγκεκριμένη συνάρτηση πιθανότητας $P : dom(X) \mapsto \Delta(dom(X))$ τ.ώ: $\forall x \in dom(X),\ P(x) = \mathbb{P}[X = x]$.

ΟΡΙΣΜΟΣ 9.1: Εντροπία

[Mitzenmacher-Upfal (2005), Chapter 9]

- Η **εντροπία σε bits** μιας διακριτής τυχαίας μεταβλητής X ορίζεται ως εξής:

$$H(X) = -\sum_{x \in dom(X)} \mathbb{P}[X = x] \log_2(\mathbb{P}[X = x]) = \mathbb{E}\left[\log_2\left(\frac{1}{P(X)}\right)\right]$$

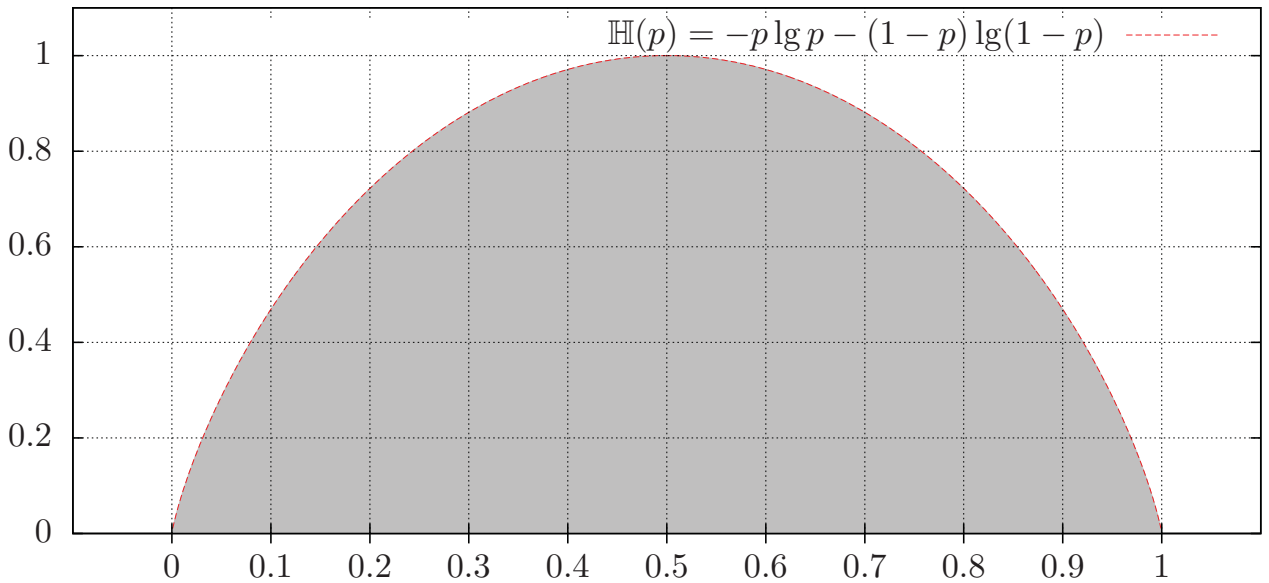
- Η **δυαδική συνάρτηση εντροπίας** μιας **δυαδικής** τυχαίας μεταβλητής X (δλδ, με $dom(X) = \{H, T\}$) όπου $\mathbb{P}[X = H] = p$ είναι:

$$\boxed{H(p) = -p \cdot \log_2(p) - (1-p) \cdot \log_2(1-p)}$$

Εντροπία (II)

- Θεωρούμε ότι $H(0) = H(1) = 0$, ώστε η $H(p)$ να είναι συνεχής στο $[0, 1]$.

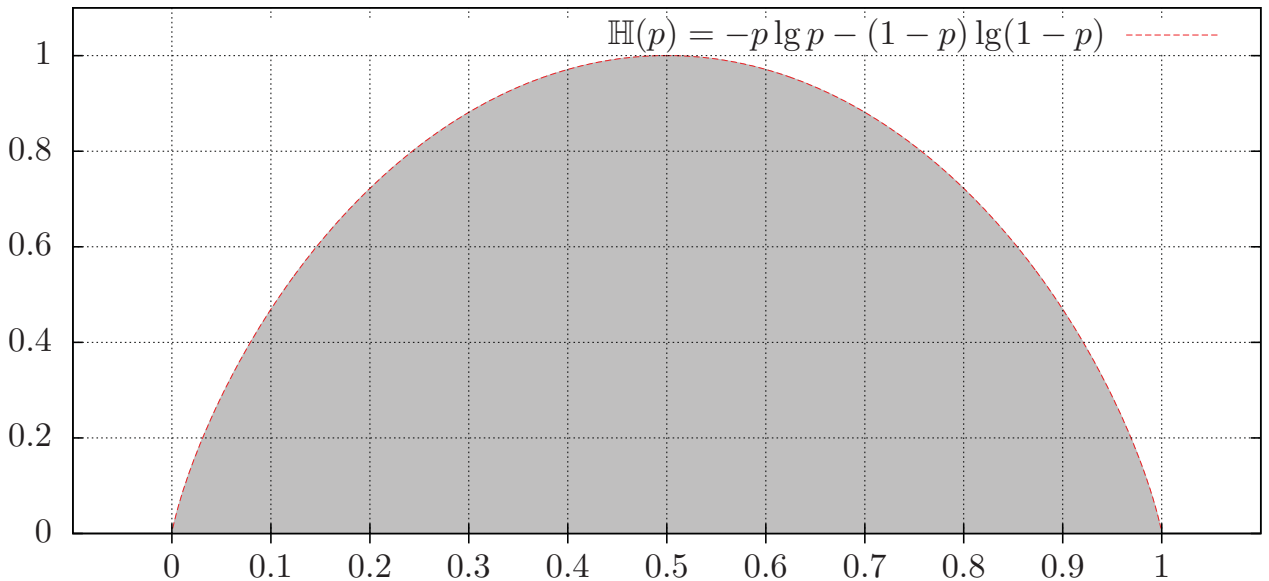
$\therefore$ Η δυαδική συνάρτηση εντροπίας είναι συνεχής στο πεδίο ορισμού $[0, 1]$:



Εντροπία (II)

- Θεωρούμε ότι $H(0) = H(1) = 0$, ώστε η $H(p)$ να είναι συνεχής στο $[0, 1]$.

$\therefore$ Η δυαδική συνάρτηση εντροπίας είναι συνεχής στο πεδίο ορισμού $[0, 1]$:

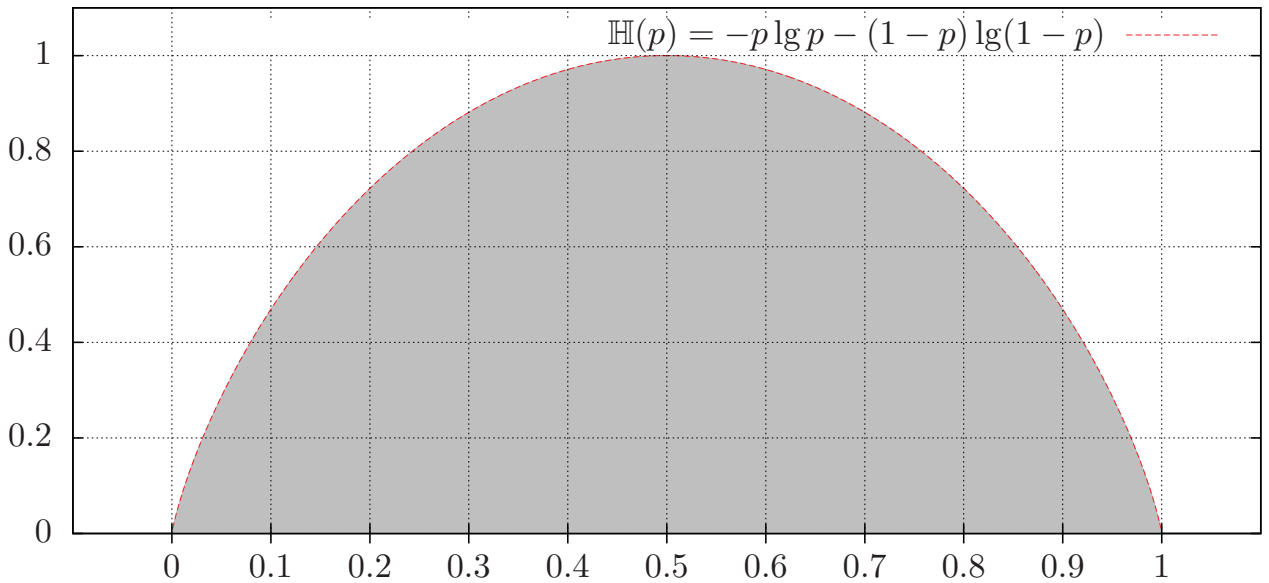


Q1 Για κέρματα $[H : 3/4\ ,\ T : 1/4]$ και $[H : 7/8\ ,\ T : 1/8]$, ποιο από τα δυο έχει μεγαλύτερη εντροπία;

Εντροπία (II)

- Θεωρούμε ότι $H(0) = H(1) = 0$, ώστε η $H(p)$ να είναι συνεχής στο $[0, 1]$.

$\therefore$ Η δυαδική συνάρτηση εντροπίας είναι συνεχής στο πεδίο ορισμού $[0, 1]$:



Q1 Για κέρματα $[H : 3/4\ ,\ T : 1/4]$ και $[H : 7/8\ ,\ T : 1/8]$, ποιο από τα δυο έχει μεγαλύτερη εντροπία;

A1 $H(3/4) \approx 0.8113$ και $H(7/8) \approx 0.5436$.

Εντροπία (III)

Q2 Για ποια τιμή του $p \in (0, 1)$ μεγιστοποιείται η τιμή της δυαδικής συνάρτησης εντροπίας;

Εντροπία (III)

Q2 Για ποια τιμή του $p \in (0,1)$ μεγιστοποιείται η τιμή της δυαδικής συνάρτησης εντροπίας;

A2 $\frac{dH(p^*)}{dp} = -\log_2\left(\frac{1-p^*}{p^*}\right) = 0 \Rightarrow \frac{1-p^*}{p^*} = 1 \Rightarrow \boxed{p^* = \frac{1}{2}}$

$$\frac{d^2H(p)}{dp^2} = -\frac{1}{p(1-p)} < 0, \ \ \forall p \in (0,1)$$

Εντροπία (III)

Q2 Για ποια τιμή του $p \in (0,1)$ μεγιστοποιείται η τιμή της δυαδικής συνάρτησης εντροπίας;

A2 $\frac{dH(p^*)}{dp} = -\log_2\left(\frac{1-p^*}{p^*}\right) = 0 \Rightarrow \frac{1-p^*}{p^*} = 1 \Rightarrow \boxed{p^* = \frac{1}{2}}$

$\frac{d^2H(p)}{dp^2} = -\frac{1}{p(1-p)} < 0, \ \forall p \in (0,1)$

Σ2 Η μέγιστη εντροπία (ανά ρίψη) επιτυγχάνεται από ένα **«δίκαιο» νόμισμα**, που *παράγει ακριβώς ένα τυχαίο ψηφίο*. Οποιοδήποτε **«μεροληπτικό» νόμισμα** παράγει ΛΙΓΟΤΕΡΟ από ένα τυχαίο ψηφίο, ανά ρίψη του.

Εντροπία (III)

Q2 Για ποια τιμή του $p \in (0,1)$ μεγιστοποιείται η τιμή της δυαδικής συνάρτησης εντροπίας;

A2 $\frac{dH(p^*)}{dp} = -\log_2\left(\frac{1-p^*}{p^*}\right) = 0 \Rightarrow \frac{1-p^*}{p^*} = 1 \Rightarrow \boxed{p^* = \frac{1}{2}}$

$\frac{d^2H(p)}{dp^2} = -\frac{1}{p(1-p)} < 0, \ \forall p \in (0,1)$

Σ2 Η μέγιστη εντροπία (ανά ρίψη) επιτυγχάνεται από ένα **«δίκαιο» νόμισμα**, που *παράγει ακριβώς ένα τυχαίο ψηφίο*. Οποιοδήποτε **«μεροληπτικό» νόμισμα** παράγει ΛΙΓΟΤΕΡΟ από ένα τυχαίο ψηφίο, ανά ρίψη του.

Q3 Ποια η εντροπία σε bits της τυχαίας μεταβλητής X_n που δηλώνει το αποτέλεσμα μιας ρίψης ενός **«δίκαιου»** ζαριού $n \geq 1$ όψεων (ή ισοδύναμα, οι n **ανεξάρτητες** ρίψεις ενός δίκαιου νομίσματος δυο όψεων);

Εντροπία (III)

Q2 Για ποια τιμή του $p \in (0,1)$ μεγιστοποιείται η τιμή της δυαδικής συνάρτησης εντροπίας;

A2 $\frac{dH(p^*)}{dp} = -\log_2\left(\frac{1-p^*}{p^*}\right) = 0 \Rightarrow \frac{1-p^*}{p^*} = 1 \Rightarrow \boxed{p^* = \frac{1}{2}}$

$\frac{d^2H(p)}{dp^2} = -\frac{1}{p(1-p)} < 0, \ \forall p \in (0,1)$

Σ2 Η μέγιστη εντροπία (ανά ρίψη) επιτυγχάνεται από ένα **«δίκαιο» νόμισμα**, που *παράγει ακριβώς ένα τυχαίο ψηφίο*. Οποιοδήποτε **«μεροληπτικό» νόμισμα** παράγει ΛΙΓΟΤΕΡΟ από ένα τυχαίο ψηφίο, ανά ρίψη του.

Q3 Ποια η εντροπία σε bits της τυχαίας μεταβλητής X_n που δηλώνει το αποτέλεσμα μιας ρίψης ενός **«δίκαιου»** ζαριού $n \geq 1$ όψεων (ή ισοδύναμα, οι n **ανεξάρτητες** ρίψεις ενός δίκαιου νομίσματος δυο όψεων);

A3 $H(X_n) = -n \cdot \frac{1}{n} \cdot \log_2\left(\frac{1}{n}\right) \Rightarrow \boxed{H(X_n) = \log_2(n)}$

Εντροπία (IV)

Q4 Τι σημαίνει ο ισχυρισμός ότι ο βαθμός τυχαιότητας που εξασφαλίζει η X_8 είναι $H(X_8) = \log_2(8) = 3$;

Εντροπία (IV)

Q4 Τι σημαίνει ο ισχυρισμός ότι ο βαθμός τυχαιότητας που εξασφαλίζει η X_8 είναι $H(X_8) = \log_2(8) = 3$;

A4 8 αποτελέσματα ζαριού αντιστοιχούν σε 8 3-ψήφιες δυαδικές αναπαραστάσεις. Κάθε αποτελεσμα του ζαριού (και η αντιστοιχη συμβολοσειρά) είναι **ισοπίθανο**. Άρα, ακριβώς τρεις ανεξάρτητες ρίψεις ενός *δίκαιου νομίσματος*.

Εντροπία (IV)

Q4 Τι σημαίνει ο ισχυρισμός ότι ο βαθμός τυχαιότητας που εξασφαλίζει η X_8 είναι $H(X_8) = \log_2(8) = 3$;

A4 8 αποτελέσματα ζαριού αντιστοιχούν σε 8 3-ψήφιες δυαδικές αναπαραστάσεις. Κάθε αποτελεσμα του ζαριού (και η αντιστοιχη συμβολοσειρά) είναι **ισοπίθανο**. Άρα, ακριβώς τρεις ανεξάρτητες ρίψεις ενός *δίκαιου νομίσματος*.

- Η εντροπία $H(X)$ μιας μεταβλητής X είναι ένα μέτρο της τυχαιότητάς της, που ΔΕΝ εξαρτάται από το πεδίο τιμών της (όπως πχ η αναμενόμενη τιμή $\mathbb{E}[X]$ ή η διακύμανση $\mathbb{V}[X]$) αλλά από τον *τρόπο ανάθεσης μαζών πιθανότητας* στις διαφορετικές τιμές που μπορεί να πάρει η X.

Εντροπία (IV)

Q4 Τι σημαίνει ο ισχυρισμός ότι ο βαθμός τυχαιότητας που εξασφαλίζει η X_8 είναι $H(X_8) = \log_2(8) = 3$;

A4 8 αποτελέσματα ζαριού αντιστοιχούν σε 8 3-ψήφιες δυαδικές αναπαραστάσεις. Κάθε αποτελεσμα του ζαριού (και η αντιστοιχη συμβολοσειρά) είναι **ισοπίθανο**. Άρα, ακριβώς τρεις ανεξάρτητες ρίψεις ενός *δίκαιου νομίσματος*.

- Η εντροπία $H(X)$ μιας μεταβλητής X είναι ένα μέτρο της τυχαιότητάς της, που ΔΕΝ εξαρτάται από το πεδίο τιμών της (όπως πχ η αναμενόμενη τιμή $\mathbb{E}[X]$ ή η διακύμανση $\mathbb{V}[X]$) αλλά από τον *τρόπο ανάθεσης μαζών πιθανότητας* στις διαφορετικές τιμές που μπορεί να πάρει η X.

ΛΗΜΜΑ 9.1 [Mitzenmacher-Upfal (2005), Chapter 9]

Έστω X_1, X_2 *ανεξάρτητες* τυχαίες μεταβλητές και $Y = (X_1, X_2)$. Τότε: $H(Y) = H(X_1) + H(X_2)$.

Εντροπία και Διωνυμικοί Συντελεστές

ΛΗΜΜΑ 9.2 [Mitzenmacher-Upfal (2005), Chapter 9]

Έστω ότι για φυσικό αριθμό $n \geq 1$ και $q \in [0, 1]$ ισχύει ότι $qn \in \mathbb{N}$. Τότε: $\boxed{\frac{2^{nH(q)}}{n+1} \leq C(n, qn) \leq 2^{nH(q)}}$

Εντροπία και Διωνυμικοί Συντελεστές

ΛΗΜΜΑ 9.2

[Mitzenmacher-Upfal (2005), Chapter 9]

Έστω ότι για φυσικό αριθμό $n \geq 1$ και $q \in [0,1]$ ισχύει ότι $qn \in \mathbb{N}$. Τότε: $\boxed{\frac{2^{nH(q)}}{n+1} \leq C(n,qn) \leq 2^{nH(q)}}$

ΠΟΡΙΣΜΑ 9.3

[Mitzenmacher-Upfal (2005), Chapter 9]

1. Για $q \in [0,1/2]$: $C(n,\lfloor qn \rfloor) \leq 2^{nH(q)}$ και $C(n,\lceil qn \rceil) \geq \frac{2^{nH(q)}}{n+1}$
2. Για $q \in [1/2,1]$: $C(n,\lceil qn \rceil) \leq 2^{nH(q)}$ και $C(n,\lfloor qn \rfloor) \geq \frac{2^{nH(q)}}{n+1}$

Εντροπία και Διωνυμικοί Συντελεστές

ΛΗΜΜΑ 9.2

[Mitzenmacher-Upfal (2005), Chapter 9]

Έστω ότι για φυσικό αριθμό $n \geq 1$ και $q \in [0,1]$ ισχύει ότι $qn \in \mathbb{N}$. Τότε: $\boxed{\frac{2^{nH(q)}}{n+1} \leq C(n,qn) \leq 2^{nH(q)}}$

ΠΟΡΙΣΜΑ 9.3

[Mitzenmacher-Upfal (2005), Chapter 9]

1. Για $q \in [0,1/2]$: $C(n,\lfloor qn \rfloor) \leq 2^{nH(q)}$ και $C(n,\lceil qn \rceil) \geq \frac{2^{nH(q)}}{n+1}$
2. Για $q \in [1/2,1]$: $C(n,\lceil qn \rceil) \leq 2^{nH(q)}$ και $C(n,\lfloor qn \rfloor) \geq \frac{2^{nH(q)}}{n+1}$

- Για n **ανεξάρτητες** ρίψεις του $[H : p > 0.5\ ,\ T : 1-p]$:
  - Από **Chernoff Bound**: $(1+o(1))np \approx np$ εμφανίσεις του H, με μεγάλη πιθανότητα.
  - $C(n,\lceil np \rceil) \in \left(\frac{1}{n+1}, 1\right) 2^{nH(p)}$ «ισοπίθανες» συμβολοσειρές.
  - Κάθε συμβολοσειρά με $\lceil np \rceil$ εμφανίσεις του H έχει πιθανότητα εμφάνισης $\approx p^{\lceil pn \rceil}(1-p)^{n-\lceil pn \rceil} \in (1, n+1)\, 2^{-nH(p)}$.

Σκελετός Ομιλίας

Συνάρτηση Εξαγωγής Τυχαιότητας

Q5 Πόσες **ανεξάρτητες ρίψεις** ενός **«δίκαιου» νομίσματος** μπορούμε να εξάγουμε (κατά μέσο όρο) από μια ανάθεση τιμής σε μια τυχαία μεταβλητή X;

Συνάρτηση Εξαγωγής Τυχαιότητας

Q5 Πόσες **ανεξάρτητες ρίψεις** ενός **«δίκαιου» νομίσματος** μπορούμε να εξάγουμε (κατά μέσο όρο) από μια ανάθεση τιμής σε μια τυχαία μεταβλητή X;

ΟΡΙΣΜΟΣ 9.2

[Mitzenmacher-Upfal (2005), Chapter 9]

Έστω $|y|$ το πλήθος των bits μιας δυαδικής συμβολοσειράς y. Μια **συνάρτηση εξαγωγής** (τυχαιότητας) $Ext(X)$ δέχεται ως είσοδο την τιμή μιας τυχαίας μεταβλητής X και επιστρέφει μια δυαδική συμβολοσειρά y τέτοια ώστε να ισχύει το εξής:

$$\forall k \in \mathbb{N},\ \mathbb{P}[|Ext(X)| = k] > 0 \ \rightarrow\ \mathbb{P}[Ext(X) = y \mid |y| = k] = 2^{-k}$$

Συνάρτηση Εξαγωγής Τυχαιότητας

Q5 Πόσες **ανεξάρτητες ρίψεις** ενός **«δίκαιου» νομίσματος** μπορούμε να εξάγουμε (κατά μέσο όρο) από μια ανάθεση τιμής σε μια τυχαία μεταβλητή X;

ΟΡΙΣΜΟΣ 9.2

[Mitzenmacher-Upfal (2005), Chapter 9]

Έστω $|y|$ το πλήθος των bits μιας δυαδικής συμβολοσειράς y. Μια **συνάρτηση εξαγωγής** (τυχαιότητας) $Ext(X)$ δέχεται ως είσοδο την τιμή μιας τυχαίας μεταβλητής X και επιστρέφει μια δυαδική συμβολοσειρά y τέτοια ώστε να ισχύει το εξής:

$$\forall k \in \mathbb{N},\ \mathbb{P}[|Ext(X)| = k] > 0 \ \rightarrow\ \mathbb{P}[Ext(X) = y \mid |y| = k] = 2^{-k}$$

ΠΑΡΑΤΗΡΗΣΗ 1: Δεν απαιτούμε η συνάρτηση Ext να είναι αποδοτικά υλοποιήσιμη (πχ, σε πολυωνυμικό χρόνο).

Συνάρτηση Εξαγωγής Τυχαιότητας

Q5 Πόσες **ανεξάρτητες ρίψεις** ενός **«δίκαιου» νομίσματος** μπορούμε να εξάγουμε (κατά μέσο όρο) από μια ανάθεση τιμής σε μια τυχαία μεταβλητή X;

ΟΡΙΣΜΟΣ 9.2

[Mitzenmacher-Upfal (2005), Chapter 9]

Έστω $|y|$ το πλήθος των bits μιας δυαδικής συμβολοσειράς y. Μια **συνάρτηση εξαγωγής** (τυχαιότητας) $Ext(X)$ δέχεται ως είσοδο την τιμή μιας τυχαίας μεταβλητής X και επιστρέφει μια δυαδική συμβολοσειρά y τέτοια ώστε να ισχύει το εξής:

$$\forall k \in \mathbb{N},\ \mathbb{P}[|Ext(X)| = k] > 0 \ \rightarrow\ \mathbb{P}[Ext(X) = y \mid |y| = k] = 2^{-k}$$

ΠΑΡΑΤΗΡΗΣΗ 1: Δεν απαιτούμε η συνάρτηση Ext να είναι αποδοτικά υλοποιήσιμη (πχ, σε πολυωνυμικό χρόνο).

ΠΑΡΑΤΗΡΗΣΗ 2: Δεν παράγονται κατ' ανάγκη συγκεκριμένου μήκους ακολουθίες στην έξοδο της Ext.

Εξαγωγή Τυχαιότητας (I)
...από Ομοιόμορφες Τυχαίες Μεταβλητές

Q6 Πόσες **ανεξάρτητες ριψεις** ενός δίκαιου νομίσματος μπορούν να εξαχθούν από μια *ομοιόμορφα κατανεμημένη* τυχαία μεταβλητή $X_8 \in_{\textbf{uar}} \{0, 1, 2, \ldots, 7\}$;

Εξαγωγή Τυχαιότητας (I)
...από Ομοιόμορφες Τυχαίες Μεταβλητές

Q6 Πόσες **ανεξάρτητες ριψεις** ενός δίκαιου νομίσματος μπορούν να εξαχθούν από μια *ομοιόμορφα κατανεμημένη* τυχαία μεταβλητή $X_8 \in_{\textbf{uar}} \{0, 1, 2, \ldots, 7\}$;

A6 $Ext(X) = (X)_2$ /* 3-ψήφια δυαδική αναπαράσταση της τιμής της X */

Εξαγωγή Τυχαιότητας (I)
...από Ομοιόμορφες Τυχαίες Μεταβλητές

Q6 Πόσες **ανεξάρτητες ριψεις** ενός δίκαιου νομίσματος μπορούν να εξαχθούν από μια *ομοιόμορφα κατανεμημένη* τυχαία μεταβλητή $X_8 \in_{\textbf{uar}} \{0, 1, 2, \ldots, 7\}$;

A6 $Ext(X) = (X)_2$ /* 3-ψήφια δυαδική αναπαράσταση της τιμής της X */

Q7 Πόσες **ανεξάρτητες ριψεις** ενός δίκαιου νομίσματος μπορούν να εξαχθούν από μια *ομοιόμορφα κατανεμημένη* τυχαία μεταβλητή $X_{12} \in_{\textbf{uar}} \{0, 1, 2, \ldots, 11\}$;

Εξαγωγή Τυχαιότητας (Ι)

...από Ομοιόμορφες Τυχαίες Μεταβλητές

Q6 Πόσες **ανεξάρτητες ριψεις** ενός δίκαιου νομίσματος μπορούν να εξαχθούν από μια *ομοιόμορφα κατανεμημένη* τυχαία μεταβλητή $X_8 \in_{\textbf{uar}} \{0, 1, 2, \ldots, 7\}$;

A6 $Ext(X) = (X)_2$ /* 3-ψήφια δυαδική αναπαράσταση της τιμής της X */

Q7 Πόσες **ανεξάρτητες ριψεις** ενός δίκαιου νομίσματος μπορούν να εξαχθούν από μια *ομοιόμορφα κατανεμημένη* τυχαία μεταβλητή $X_{12} \in_{\textbf{uar}} \{0, 1, 2, \ldots, 11\}$;

A7 $$Ext(X) = \begin{cases} (X)_2, & X \leq 7 \quad \text{/* 3-ψήφια δυαδική συμβολοσειρά */} \\ (X-8)_2, & X \geq 8 \quad \text{/* 2-ψήφια δυαδική συμβολοσειρά */} \end{cases}$$

Εξαγωγή Τυχαιότητας (I)

...από Ομοιόμορφες Τυχαίες Μεταβλητές

Q6 Πόσες **ανεξάρτητες ριψεις** ενός δίκαιου νομίσματος μπορούν να εξαχθούν από μια *ομοιόμορφα κατανεμημένη* τυχαία μεταβλητή $X_8 \in_{\mathbf{uar}} \{0, 1, 2, \ldots, 7\}$;

A6 $Ext(X) = (X)_2$ /* 3-ψήφια δυαδική αναπαράσταση της τιμής της X */

Q7 Πόσες **ανεξάρτητες ριψεις** ενός δίκαιου νομίσματος μπορούν να εξαχθούν από μια *ομοιόμορφα κατανεμημένη* τυχαία μεταβλητή $X_{12} \in_{\mathbf{uar}} \{0, 1, 2, \ldots, 11\}$;

A7 $$Ext(X) = \begin{cases} (X)_2, & X \leq 7 \quad \text{/* 3-ψήφια δυαδική συμβολοσειρά */} \\ (X-8)_2, & X \geq 8 \quad \text{/* 2-ψήφια δυαδική συμβολοσειρά */} \end{cases}$$

ΘΕΩΡΗΜΑ 9.4 [Mitzenmacher-Upfal (2005), Chapter 9]

Για $X \in_{\mathbf{uar}} \{0, 1, 2, \ldots, m-1\}$, υπάρχει συνάρτηση εξαγωγής $Ext(X)$ που παρέχει *κατά μέσο όρο* τουλάχιστον $\lfloor \log_2(m) \rfloor - 1 = \lfloor H(X) \rfloor - 1$ **ανεξάρτητες ρίψεις** δίκαιου νομίσματος.

Εξαγωγή Τυχαιότητας (II)
...από Ομοιόμορφες Τυχαίες Μεταβλητές

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.4)

1. Κατασκευή:

$$Ext(X) = \begin{cases} \epsilon, & m < 2 \\ (X)_2, & m = 2^a \geq 2 \\ (X)_2, & 2 \leq 2^a < m < 2^{a+1},\ X \leq 2^a - 1 \\ Ext(X - 2^a) & 2 \leq 2^a < m < 2^{a+1},\ X \geq 2^a \end{cases}$$

/* κενή συμβολοσειρά */
/* $|Ext(X)| = a$ */
/* $|Ext(X)| = a$ */
/* $|Ext(X)| \leq a - 1$ */

Εξαγωγή Τυχαιότητας (II)
...από Ομοιόμορφες Τυχαίες Μεταβλητές

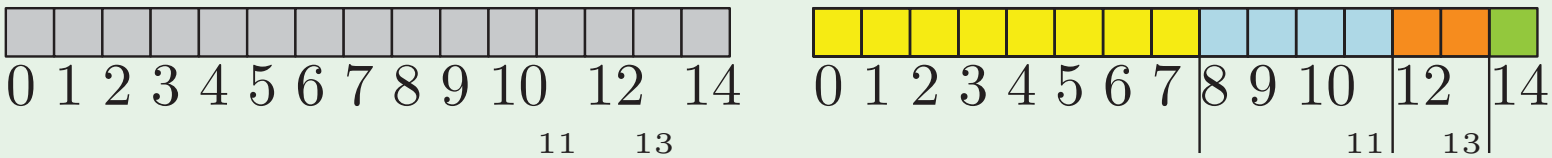
ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.4)

1. Κατασκευή:

$$Ext(X) = \begin{cases} \epsilon, & m < 2 \\ (X)_2, & m = 2^a \geq 2 \\ (X)_2, & 2 \leq 2^a < m < 2^{a+1},\ X \leq 2^a - 1 \\ Ext(X - 2^a) & 2 \leq 2^a < m < 2^{a+1},\ X \geq 2^a \end{cases}$$

/* κενή συμβολοσειρά */
/* $|Ext(X)| = a$ */
/* $|Ext(X)| = a$ */
/* $|Ext(X)| \leq a - 1$ */

ΠΑΡΑΔΕΙΓΜΑ: Για $m = 15$ και $X = 10$ έχουμε:



Εξαγωγή Τυχαιότητας (II)

...από Ομοιόμορφες Τυχαίες Μεταβλητές

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.4)

1. Κατασκευή:

$$Ext(X) = \begin{cases} \epsilon, & m < 2 \\ (X)_2, & m = 2^a \geq 2 \\ (X)_2, & 2 \leq 2^a < m < 2^{a+1},\ X \leq 2^a - 1 \\ Ext(X - 2^a) & 2 \leq 2^a < m < 2^{a+1},\ X \geq 2^a \end{cases}$$

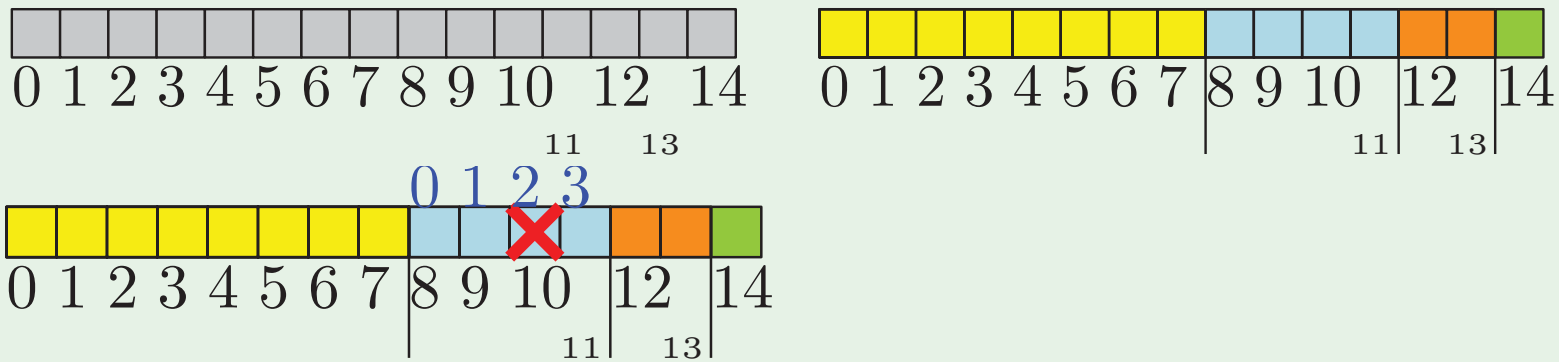
/* κενή συμβολοσειρά */

/* $|Ext(X)| = a$ */

/* $|Ext(X)| = a$ */

/* $|Ext(X)| \leq a - 1$ */

ΠΑΡΑΔΕΙΓΜΑ: Για $m = 15$ και $X = 10$ έχουμε:



Εξαγωγή Τυχαιότητας (II)

...από Ομοιόμορφες Τυχαίες Μεταβλητές

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.4)

1. Κατασκευή:

$$Ext(X) = \begin{cases} \epsilon, & m < 2 \\ (X)_2, & m = 2^a \geq 2 \\ (X)_2, & 2 \leq 2^a < m < 2^{a+1},\ X \leq 2^a - 1 \\ Ext(X - 2^a) & 2 \leq 2^a < m < 2^{a+1},\ X \geq 2^a \end{cases}$$

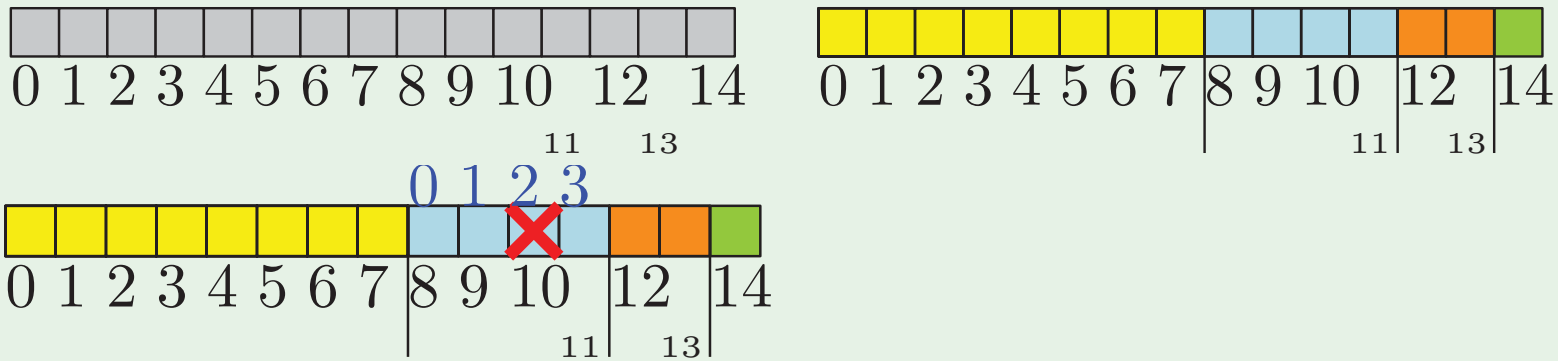
/* κενή συμβολοσειρά */

/* $|Ext(X)| = a$ */

/* $|Ext(X)| = a$ */

/* $|Ext(X)| \leq a - 1$ */

ΠΑΡΑΔΕΙΓΜΑ: Για $m = 15$ και $X = 10$ έχουμε:



$$Ext(X) = (10 - 2^3)_2 = 10$$

Εξαγωγή Τυχαιότητας (III)

...από Ομοιόμορφες Τυχαίες Μεταβλητές

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.4)

2. Ορθότητα: Κάθε k-ψηφια συμβολοσειρά που επιστρέφεται (για $k \geq 1$) έχει πιθανότητα εμφάνισης 2^{-k}.

Εξαγωγή Τυχαιότητας (III)

...από Ομοιόμορφες Τυχαίες Μεταβλητές

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.4)

2. Ορθότητα: Κάθε k-ψηφια συμβολοσειρά που επιστρέφεται (για $k \geq 1$) έχει πιθανότητα εμφάνισης 2^{-k}.
3. Αναμενόμενο μήκος εξόδου:
   - Για $m = 1$ επιστρέφονται 0 τυχαία bits.

Εξαγωγή Τυχαιότητας (III)

...από Ομοιόμορφες Τυχαίες Μεταβλητές

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.4)

2. Ορθότητα: Κάθε k-ψηφια συμβολοσειρά που επιστρέφεται (για $k \geq 1$) έχει πιθανότητα εμφάνισης 2^{-k}.
3. Αναμενόμενο μήκος εξόδου:
   - Για $m = 1$ επιστρέφονται 0 τυχαία bits.
   - Για $m = 2^a \geq 2$ επιστρέφονται $a = \log_2(m)$ τυχαία bits.

Εξαγωγή Τυχαιότητας (III)

...από Ομοιόμορφες Τυχαίες Μεταβλητές

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.4)

2. Ορθότητα: Κάθε k-ψηφια συμβολοσειρά που επιστρέφεται (για $k \geq 1$) έχει πιθανότητα εμφάνισης 2^{-k}.
3. Αναμενόμενο μήκος εξόδου:
   - Για $m = 1$ επιστρέφονται 0 τυχαία bits.
   - Για $m = 2^a \geq 2$ επιστρέφονται $a = \log_2(m)$ τυχαία bits.
   - Για $2 \leq 2^a < m < 2^{a+1}$ επιστρέφονται τουλάχιστον

$$
\begin{aligned}
& a \cdot \frac{2^a}{m} + \left(\underbrace{\lfloor \log_2 (m - 2^a) \rfloor}_{=\beta \in \{0,1,\ldots,a-1\}} - 1 \right) \cdot \left(1 - \frac{2^a}{m} \right) \\
= \;& a + (\beta - 1 - a) \cdot \left(1 - \frac{2^a}{m} \right) \\
\geq \;& a + (\beta - 1 - a) \cdot \left(1 - \frac{2^a}{2^a + 2^\beta} \right) \; /* \; \tfrac{d}{dm}\left(\tfrac{m-2^a}{m}\right) = \tfrac{2^a}{m^2} \Rightarrow \tfrac{m-2^a}{m} \leq \tfrac{2^\beta}{2^a+2^\beta} \; */ \\
= \;& a + (\beta - 1 - a) \cdot \tfrac{2^\beta}{2^a + 2^\beta} \geq a - 1 = \lfloor \log_2(m) \rfloor - 1
\end{aligned}
$$

   τυχαία bits.

Εξαγωγή Τυχαιότητας (III)

...από Ομοιόμορφες Τυχαίες Μεταβλητές

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.4)

2. Ορθότητα: Κάθε k-ψηφια συμβολοσειρά που επιστρέφεται (για $k \geq 1$) έχει πιθανότητα εμφάνισης 2^{-k}.
3. Αναμενόμενο μήκος εξόδου:
   - Για $m = 1$ επιστρέφονται 0 τυχαία bits.
   - Για $m = 2^a \geq 2$ επιστρέφονται $a = \log_2(m)$ τυχαία bits.
   - Για $2 \leq 2^a < m < 2^{a+1}$ επιστρέφονται τουλάχιστον

$$
\begin{aligned}
& a \cdot \frac{2^a}{m} + \left(\underbrace{\lfloor \log_2 (m - 2^a) \rfloor}_{=\beta \in \{0,1,\ldots,a-1\}} - 1 \right) \cdot \left(1 - \frac{2^a}{m} \right) \\
= \;& a + (\beta - 1 - a) \cdot \left(1 - \frac{2^a}{m} \right) \\
\geq \;& a + (\beta - 1 - a) \cdot \left(1 - \frac{2^a}{2^a + 2^\beta} \right) \; /* \; \tfrac{d}{dm}\left(\tfrac{m-2^a}{m}\right) = \tfrac{2^a}{m^2} \Rightarrow \tfrac{m-2^a}{m} \leq \tfrac{2^\beta}{2^a+2^\beta} \; */ \\
= \;& a + (\beta - 1 - a) \cdot \frac{2^\beta}{2^a + 2^\beta} \geq a - 1 = \lfloor \log_2(m) \rfloor - 1
\end{aligned}
$$

τυχαία bits. ■

Εξαγωγή Τυχαιότητας (I)

...από «Πειραγμένα» Νομίσματα

ΘΕΩΡΗΜΑ 9.5 [Mitzenmacher-Upfal (2005), Chapter 9]

Έστω «πειραγμένο» νόμισμα $COIN(p)$ που έχει $\mathbb{P}[HEADS] = p > \frac{1}{2}$. $\forall \delta > 0$, και για αρκούντως μεγάλο $n \in \mathbb{N}$, δίνεται ως είσοδος το αποτέλεσμα n **ανεξάρτητων ρίψεων** του $COIN(p)$.

(α) Υπάρχει συνάρτηση εξαγωγής που επιστρέφει (κατά μέσο όρο) τουλάχιστον $(1-\delta) \cdot n \cdot H(p)$ **ανεξάρτητες ρίψεις** ενός δίκαιου νομίσματος $COIN(1/2)$.

(β) Ο αναμενόμενος αριθμός ανεξάρτητων ρίψεων ενός δίκαιου νομίσματος $COIN(1/2)$ που μπορεί να επιστρέψει οποιαδηποτε συνάρτηση εξαγωγής είναι το πολύ $n \cdot H(p)$.

Εξαγωγή Τυχαιότητας (II)

...από «Πειραγμένα» Νομίσματα

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.5)

1. Κατασκευή:
   - $C(n,k)$ $n-$ψήφιες δυαδικές συμβολοσειρές με **ακριβώς** k $HEADS$, καθεμιά με πιθανότητα εμφάνισης $p^k \cdot (1-p)^{n-k}$.
   - Για κάθε συμβολοσειρά σ_k με ακριβώς k εμφανίσεις του $HEADS$, $X(\sigma_k) =$ **μοναδικός** αριθμός από το $S_k = \{0, 1, 2, \ldots, C(n,k) - 1\}$.
   - Επιστρέφεται η συμβολοσειρά $Ext(X(\sigma_k))$.

Εξαγωγή Τυχαιότητας (II)

...από «Πειραγμένα» Νομίσματα

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.5)

1. Κατασκευή:
   - $C(n,k)$ $n-$ψήφιες δυαδικές συμβολοσειρές με **ακριβώς** k *HEADS*, καθεμιά με πιθανότητα εμφάνισης $p^k \cdot (1-p)^{n-k}$.
   - Για κάθε συμβολοσειρά σ_k με ακριβώς k εμφανίσεις του *HEADS*, $X(\sigma_k) =$ **μοναδικός** αριθμός από το $S_k = \{0, 1, 2, \ldots, C(n,k) - 1\}$.
   - Επιστρέφεται η συμβολοσειρά $Ext(X(\sigma_k))$.
2. Ορθότητα: Για οποιοδήποτε (συγκεκριμένο) πλήθος k εμφανίσεων *HEADS* μεταξύ των n ανεξάρτητων ρίψεων του $COIN(p)$, επιστρέφονται τουλάχιστον $\lfloor \log_2 (C(n,k)) \rfloor - 1$ ανεξάρτητες ρίψεις του $COIN(1/2)$.

Εξαγωγή Τυχαιότητας (III)

...από «Πειραγμένα» Νομίσματα

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.5)

3. Κάτω φράγμα για πλήθος B των τυχαίων bits:

Εξαγωγή Τυχαιότητας (III)

...από «Πειραγμένα» Νομίσματα

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.5)

3. Κάτω φράγμα για πλήθος B των τυχαίων bits:

$$\begin{aligned}
\mathbb{E}[B] &\geq \sum_{k=0}^{n} \mathbb{P}[\#H = k] \cdot (\lfloor \log_2(C(n,k)) \rfloor - 1) \\
&\geq \sum_{k=\lfloor n(p-\epsilon) \rfloor}^{\lceil n(p+\epsilon) \rceil} \mathbb{P}[\#H = k] \cdot (\lfloor \log_2(C(n,k)) \rfloor - 1) \\
&\geq \sum_{k=\lfloor n(p-\epsilon) \rfloor}^{\lceil n(p+\epsilon) \rceil} \mathbb{P}[\#H = k] \cdot \left(\left\lfloor \log_2 \left(\frac{2^{nH(p+\epsilon)}}{n+1} \right) \right\rfloor - 1 \right) \\
&\geq (nH(p+\epsilon) - \log_2(n+1) - 2) \cdot \sum_{k=\lfloor n(p-\epsilon) \rfloor}^{\lceil n(p+\epsilon) \rceil} \mathbb{P}[\#H = k]
\end{aligned}$$

Εξαγωγή Τυχαιότητας (III)

...από «Πειραγμένα» Νομίσματα

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.5)

3. Κάτω φράγμα για πλήθος B των τυχαίων bits:

$$\begin{aligned}
\mathbb{E}[B] &\geq \sum_{k=0}^{n} \mathbb{P}[\#H = k] \cdot (\lfloor \log_2(C(n,k)) \rfloor - 1) \\
&\geq \sum_{k=\lfloor n(p-\epsilon) \rfloor}^{\lceil n(p+\epsilon) \rceil} \mathbb{P}[\#H = k] \cdot (\lfloor \log_2(C(n,k)) \rfloor - 1) \\
&\geq \sum_{k=\lfloor n(p-\epsilon) \rfloor}^{\lceil n(p+\epsilon) \rceil} \mathbb{P}[\#H = k] \cdot \left(\left\lfloor \log_2 \left(\frac{2^{nH(p+\epsilon)}}{n+1} \right) \right\rfloor - 1 \right) \\
&\geq (nH(p+\epsilon) - \log_2(n+1) - 2) \cdot \sum_{k=\lfloor n(p-\epsilon) \rfloor}^{\lceil n(p+\epsilon) \rceil} \mathbb{P}[\#H = k]
\end{aligned}$$

CHERNOFF BOUND: $\mathbb{P}[|\#H - pn| > \epsilon n] \leq 2 \exp\left(-\frac{pn\epsilon^2}{3}\right)$

Εξαγωγή Τυχαιότητας (III)

...από «Πειραγμένα» Νομίσματα

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.5)

3. Κάτω φράγμα για πλήθος B των τυχαίων bits:

$$\begin{aligned}
\mathbb{E}[B] &\geq \sum_{k=0}^{n} \mathbb{P}[\#H = k] \cdot (\lfloor \log_2(C(n,k)) \rfloor - 1) \\
&\geq \sum_{k=\lfloor n(p-\epsilon) \rfloor}^{\lceil n(p+\epsilon) \rceil} \mathbb{P}[\#H = k] \cdot (\lfloor \log_2(C(n,k)) \rfloor - 1) \\
&\geq \sum_{k=\lfloor n(p-\epsilon) \rfloor}^{\lceil n(p+\epsilon) \rceil} \mathbb{P}[\#H = k] \cdot \left(\left\lfloor \log_2 \left(\frac{2^{nH(p+\epsilon)}}{n+1} \right) \right\rfloor - 1 \right) \\
&\geq (nH(p+\epsilon) - \log_2(n+1) - 2) \cdot \sum_{k=\lfloor n(p-\epsilon) \rfloor}^{\lceil n(p+\epsilon) \rceil} \mathbb{P}[\#H = k]
\end{aligned}$$

CHERNOFF BOUND: $\mathbb{P}[|\#H - pn| > \epsilon n] \leq 2 \exp\left(-\frac{pn\epsilon^2}{3}\right)$

$$\begin{aligned}
\mathbb{E}[B] &\geq (nH(p+\epsilon) - \log_2(n+1) - 2) \cdot \left[1 - 2\exp\left(-\frac{pn\epsilon^2}{3}\right)\right] \\
&\geq \left(1 - \frac{2\delta}{3}\right) nH(p) \cdot \left(1 - \frac{\delta}{3}\right) \quad /* \text{ με κατάλληλη επιλογή } \epsilon > 0 \text{ και } n \in \mathbb{N} */ \\
&> (1-\delta) nH(p)
\end{aligned}$$

Εξαγωγή Τυχαιότητας (IV)

...από «Πειραγμένα» Νομίσματα

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.5)

4. Άνω φράγμα στο πλήθος B των τυχαίων bits:
   - $\Sigma =$ τυχαία συμβολοσειρά από τις n ρίψεις του $COIN(p)$.
   - $Ext =$ αυθαιρετη συνάρτηση εξαγωγής, με είσοδο την Σ.

Εξαγωγή Τυχαιότητας (IV)

...από «Πειραγμένα» Νομίσματα

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.5)

4. Άνω φράγμα στο πλήθος B των τυχαίων bits:
   - $\Sigma =$ τυχαία συμβολοσειρά από τις n ρίψεις του $COIN(p)$.
   - $Ext =$ αυθαιρετη συνάρτηση εξαγωγής, με είσοδο την Σ.
   - ΙΣΧΥΡΙΣΜΟΣ: $\forall \sigma \in \{0,1\}^n,\ |Ext(\sigma)| \leq \log_2\left(\frac{1}{\mathbb{P}[\Sigma=\sigma]}\right)$.
     - ⋆ Όλες οι $2^{|Ext(\sigma)|}$ **συμβολοσειρές εξόδου** με μήκος $k = |Ext(\sigma)|$ είναι *ισοπίθανες*.

Εξαγωγή Τυχαιότητας (IV)

...από «Πειραγμένα» Νομίσματα

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.5)

4. Άνω φράγμα στο πλήθος B των τυχαίων bits:
   - $\Sigma =$ τυχαία συμβολοσειρά από τις n ρίψεις του $COIN(p)$.
   - $Ext =$ αυθαιρετη συνάρτηση εξαγωγής, με είσοδο την Σ.
   - ΙΣΧΥΡΙΣΜΟΣ: $\forall \sigma \in \{0,1\}^n,\ |Ext(\sigma)| \leq \log_2\left(\frac{1}{\mathbb{P}[\Sigma=\sigma]}\right)$.
     - ★ Όλες οι $2^{|Ext(\sigma)|}$ **συμβολοσειρές εξόδου** με μήκος $k = |Ext(\sigma)|$ είναι *ισοπίθανες*.
     - ★ $2^{|Ext(\sigma)|} \cdot \mathbb{P}[\Sigma = \sigma] \leq 1 \Rightarrow |Ext(\sigma)| \leq \log_2\left(\frac{1}{\mathbb{P}[\Sigma=\sigma]}\right)$

Εξαγωγή Τυχαιότητας (IV)

...από «Πειραγμένα» Νομίσματα

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.5)

4. Άνω φράγμα στο πλήθος B των τυχαίων bits:
   - $\Sigma =$ τυχαία συμβολοσειρά από τις n ρίψεις του $COIN(p)$.
   - $Ext =$ αυθαιρετη συνάρτηση εξαγωγής, με είσοδο την Σ.
   - ΙΣΧΥΡΙΣΜΟΣ: $\forall \sigma \in \{0,1\}^n,\ |Ext(\sigma)| \leq \log_2\left(\frac{1}{\mathbb{P}[\Sigma=\sigma]}\right)$.
     - ★ Όλες οι $2^{|Ext(\sigma)|}$ **συμβολοσειρές εξόδου** με μήκος $k = |Ext(\sigma)|$ είναι *ισοπίθανες*.
     - ★ $2^{|Ext(\sigma)|} \cdot \mathbb{P}[\Sigma = \sigma] \leq 1 \Rightarrow |Ext(\sigma)| \leq \log_2\left(\frac{1}{\mathbb{P}[\Sigma=\sigma]}\right)$

$$\begin{aligned} \mathbb{E}[B] &= \sum_\sigma \mathbb{P}[\Sigma=\sigma] \cdot |Ext(\sigma)| \\ \therefore \quad &\leq \sum_\sigma \mathbb{P}[\Sigma=\sigma] \log_2\left(\frac{1}{\mathbb{P}[\Sigma=\sigma]}\right) \\ &= H(\Sigma) = n \cdot H(p) \end{aligned}$$

Εξαγωγή Τυχαιότητας (IV)

...από «Πειραγμένα» Νομίσματα

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.5)

4. Άνω φράγμα στο πλήθος B των τυχαίων bits:
   - $\Sigma =$ τυχαία συμβολοσειρά από τις n ρίψεις του $COIN(p)$.
   - $Ext =$ αυθαιρετη συνάρτηση εξαγωγής, με είσοδο την Σ.
   - ΙΣΧΥΡΙΣΜΟΣ: $\forall \sigma \in \{0,1\}^n,\ |Ext(\sigma)| \leq \log_2\left(\frac{1}{\mathbb{P}[\Sigma=\sigma]}\right)$.
     - ★ Όλες οι $2^{|Ext(\sigma)|}$ **συμβολοσειρές εξόδου** με μήκος $k = |Ext(\sigma)|$ είναι *ισοπίθανες*.
     - ★ $2^{|Ext(\sigma)|} \cdot \mathbb{P}[\Sigma = \sigma] \leq 1 \Rightarrow |Ext(\sigma)| \leq \log_2\left(\frac{1}{\mathbb{P}[\Sigma=\sigma]}\right)$

$$\begin{aligned} \mathbb{E}[B] &= \sum_\sigma \mathbb{P}[\Sigma=\sigma] \cdot |Ext(\sigma)| \\ \therefore \quad &\leq \sum_\sigma \mathbb{P}[\Sigma=\sigma] \log_2\left(\frac{1}{\mathbb{P}[\Sigma=\sigma]}\right) \\ &= H(\Sigma) = n \cdot H(p) \end{aligned}$$

■

Εξαγωγή Τυχαιότητας από Δίκαιο Νόμισμα

Q Από (δεδομένο) δίκαιο νόμισμα $COIN(1/2)$, πώς μπορώ να φτιάξω ένα «πειραγμένο» νόμισμα $COIN(p)$ για $p = 2^{-a}$;

Σκελετός Ομιλίας

Αξιοποίηση Εντροπίας στη Συμπίεση (I)

ΕΙΣΟΔΟΣ: Νόμισμα $COIN(p)$ με $\mathbb{P}[HEADS] = p > 1/2$.

ΠΕΙΡΑΜΑ: n ανεξάρτητες ρίψεις του νομίσματος.

ΕΞΟΔΟΣ: Συμπαγής περιγραφή του αποτελέσματος του πειράματος.

Αξιοποίηση Εντροπίας στη Συμπίεση (I)

ΕΙΣΟΔΟΣ: Νόμισμα $COIN(p)$ με $\mathbb{P}[HEADS] = p > 1/2$.

ΠΕΙΡΑΜΑ: n ανεξάρτητες ρίψεις του νομίσματος.

ΕΞΟΔΟΣ: Συμπαγής περιγραφή του αποτελέσματος του πειράματος.

ΑΠΛΟΣ ΤΡΟΠΟΣ: Χρήση μιας n-ψήφιας δυαδικής συμβολοσειράς.

Αξιοποίηση Εντροπίας στη Συμπίεση (I)

ΕΙΣΟΔΟΣ: Νόμισμα $COIN(p)$ με $\mathbb{P}[HEADS] = p > 1/2$.

ΠΕΙΡΑΜΑ: n ανεξάρτητες ρίψεις του νομίσματος.

ΕΞΟΔΟΣ: Συμπαγής περιγραφή του αποτελέσματος του πειράματος.

ΑΠΛΟΣ ΤΡΟΠΟΣ: Χρήση μιας n-ψήφιας δυαδικής συμβολοσειράς.

ΣΤΟΧΟΣ: Εξασφάλιση του **μικρότερου δυνατού** χώρου αποθήκευσης, δίχως απώλεια πληροφορίας.

Αξιοποίηση Εντροπίας στη Συμπίεση (II)

ΠΑΡΑΔΕΙΓΜΑ

Για $p = 3/4$ και $n = 6$ ανεξάρτητες ρίψεις του $COIN(p)$:

- Προφανής λύση: 6−ψήφια δυαδική συμβολοσειρά, όπου κάθε ψηφίο καταχωρεί το αποτέλεσμα μιας διαφορετικής ρίψης του νομίσματος.

Αξιοποίηση Εντροπίας στη Συμπίεση (II)

ΠΑΡΑΔΕΙΓΜΑ

Για $p = 3/4$ και $n = 6$ ανεξάρτητες ρίψεις του $COIN(p)$:

- Προφανής λύση: 6−ψήφια δυαδική συμβολοσειρά, όπου κάθε ψηφίο καταχωρεί το αποτέλεσμα μιας διαφορετικής ρίψης του νομίσματος.
- Αποτύπωση εξόδου μέσω **συμπίεσης δίχως απώλειες**:

$$Code = \boxed{HH \mapsto 0} \boxed{HT \mapsto 10} \boxed{TH \mapsto 110} \boxed{TT \mapsto 111}$$

Αξιοποίηση Εντροπίας στη Συμπίεση (II)

ΠΑΡΑΔΕΙΓΜΑ

Για $p = 3/4$ και $n = 6$ ανεξάρτητες ρίψεις του $COIN(p)$:

- Προφανής λύση: 6−ψήφια δυαδική συμβολοσειρά, όπου κάθε ψηφίο καταχωρεί το αποτέλεσμα μιας διαφορετικής ρίψης του νομίσματος.
- Αποτύπωση εξόδου μέσω **συμπίεσης δίχως απώλειες**:

$$Code = \boxed{HH \mapsto 0} \boxed{HT \mapsto 10} \boxed{TH \mapsto 110} \boxed{TT \mapsto 111}$$

- Για το αποτέλεσμα (X, Y) δυο **ανεξάρτητων** ρίψεων, συμβολίζουμε με $|(X, Y)|$ το μήκος της παραγόμενης δυαδικής συμβολοσειράς. Τότε:

$$\mathbb{E}[|(X, Y)|] = \frac{9}{16} \cdot 1 + \frac{3}{16} \cdot 2 + \frac{3}{16} \cdot 3 + \frac{1}{16} \cdot 3 = \frac{27}{16} < 2$$

Αξιοποίηση Εντροπίας στη Συμπίεση (II)

ΠΑΡΑΔΕΙΓΜΑ

Για $p = 3/4$ και $n = 6$ ανεξάρτητες ρίψεις του $COIN(p)$:

- Προφανής λύση: 6−ψήφια δυαδική συμβολοσειρά, όπου κάθε ψηφίο καταχωρεί το αποτέλεσμα μιας διαφορετικής ρίψης του νομίσματος.
- Αποτύπωση εξόδου μέσω **συμπίεσης δίχως απώλειες**:

$$Code = \boxed{HH \mapsto 0} \boxed{HT \mapsto 10} \boxed{TH \mapsto 110} \boxed{TT \mapsto 111}$$

- Για το αποτέλεσμα (X, Y) δυο **ανεξάρτητων** ρίψεων, συμβολίζουμε με $|(X, Y)|$ το μήκος της παραγόμενης δυαδικής συμβολοσειράς. Τότε:

$$\mathbb{E}[|(X, Y)|] = \frac{9}{16} \cdot 1 + \frac{3}{16} \cdot 2 + \frac{3}{16} \cdot 3 + \frac{1}{16} \cdot 3 = \frac{27}{16} < 2$$

Σ Είναι δυνατή η συμπίεση του αποτελέσματος των n ρίψεων ενός «πειραγμένου» νομίσματος $COIN(p)$ με *λιγότερο από ένα ψηφίο ανά ρίψη* (κατά μέσο όρο).

Κώδικες Προθέματος

Κώδικες Προθέματος – Prefix Codes: Η αποκωδικοποίηση επιτυγχάνεται απλά αναλύοντας το κωδικοποιημένο μήνυμα *από αριστερά προς τα δεξιά*.

- Βασικό Χαρακτηριστικό: Η συμβολοσειρά κωδικοποίησης οποιασδήποτε λέξης **δεν είναι πρόθεμα** της συμβολοσειράς κωδικοποίησης κάποιας άλλης λέξης. Πχ:

$Code =$

$HH \mapsto 0$	$HT \mapsto 10$	$TH \mapsto 110$	$TT \mapsto 111$

Κώδικες Προθέματος

Κώδικες Προθέματος – Prefix Codes: Η αποκωδικοποίηση επιτυγχάνεται απλά αναλύοντας το κωδικοποιημένο μήνυμα *από αριστερά προς τα δεξιά*.

- Βασικό Χαρακτηριστικό: Η συμβολοσειρά κωδικοποίησης οποιασδήποτε λέξης **δεν είναι πρόθεμα** της συμβολοσειράς κωδικοποίησης κάποιας άλλης λέξης. Πχ:

  $Code =$ | $HH \mapsto 0$ | $HT \mapsto 10$ | $TH \mapsto 110$ | $TT \mapsto 111$ |

- ΠΑΡΑΔΕΙΓΜΑ: Για την παραπάνω κωδικοποίηση προθέματος έχουμε τις εξής αποκωδικοποιήσεις:
  - $011110 \mapsto HHTTHT$
  - $101100 \mapsto HTTHHH$

Κώδικες Προθέματος

Κώδικες Προθέματος – Prefix Codes: Η αποκωδικοποίηση επιτυγχάνεται απλά αναλύοντας το κωδικοποιημένο μήνυμα *από αριστερά προς τα δεξιά*.

- Βασικό Χαρακτηριστικό: Η συμβολοσειρά κωδικοποίησης οποιασδήποτε λέξης **δεν είναι πρόθεμα** της συμβολοσειράς κωδικοποίησης κάποιας άλλης λέξης. Πχ:

  $Code =$ | $HH \mapsto 0$ | $HT \mapsto 10$ | $TH \mapsto 110$ | $TT \mapsto 111$ |

- ΠΑΡΑΔΕΙΓΜΑ: Για την παραπάνω κωδικοποίηση προθέματος έχουμε τις εξής αποκωδικοποιήσεις:
  - $011110 \mapsto HHTTHT$
  - $101100 \mapsto HTTHHH$
- Δεν είναι έγκυρες όλες οι δυαδικές συμβολοσειρές:
  - $101101 \mapsto HTTH??$

Συνάρτηση Συμπίεσης

ΟΡΙΣΜΟΣ 9.3

[Mitzenmacher-Upfal (2005), Chapter 9]

Μια **συνάρτηση συμπίεσης** *Com* δέχεται ως είσοδο τη n-ψήφια συμβολοσειρά $\sigma \in \{H,T\}^n$ που αποτυπώνει το αποτέλεσμα n ανεξάρτητων ρίψεων του $COIN(p)$ με $p = \mathbb{P}[H] > 1/2$, και επιστρέφει μια **δυαδική συμβολοσειρά** τέτοια ώστε κάθε αποτέλεσμα $\sigma \in \{H,T\}^n$ του πειράματος να αντιστοιχεί σε μια *μοναδική* δυαδική συμβολοσειρά στην έξοδο.

Συνάρτηση Συμπίεσης

ΟΡΙΣΜΟΣ 9.3

[Mitzenmacher-Upfal (2005), Chapter 9]

Μια **συνάρτηση συμπίεσης** *Com* δέχεται ως είσοδο τη n-ψήφια συμβολοσειρά $\sigma \in \{H,T\}^n$ που αποτυπώνει το αποτέλεσμα n ανεξάρτητων ρίψεων του $COIN(p)$ με $p = \mathbb{P}[H] > 1/2$, και επιστρέφει μια **δυαδική συμβολοσειρά** τέτοια ώστε κάθε αποτέλεσμα $\sigma \in \{H,T\}^n$ του πειράματος να αντιστοιχεί σε μια *μοναδική* δυαδική συμβολοσειρά στην έξοδο.

- Ειδικές (πιο απαιτητικές) περιπτώσεις συναρτήσεων συμπίεσης είναι, πχ, οι κώδικες προθέματος / επιθέματος (prefix / suffix codes).
- Ο πιο ασθενής ορισμός για την συνάρτηση συμπίεσης *Com* δίνει τη δυνατότητα παροχής ισχυρότερων κάτω-φραγμάτων.

Συσχέτιση Εντροπίας και Συμπίεσης (I)

ΘΕΩΡΗΜΑ 9.6 [Mitzenmacher-Upfal (2005), Chapter 9]

Έστω «πειραγμένο» νόμισμα $COIN(p)$ με $\mathbb{P}[H] = p > 1/2$. $\forall \delta > 0$ και αρκούντως μεγάλο $n \in \mathbb{N}$, ως προς το αποτέλεσμα $\sigma \in \{H,T\}^n$ των n ανεξάρτητων ρίψεων του $COIN(p)$:

(α) Υπάρχει συνάρτηση συμπίεσης Com^* τ.ώ. με αναμενόμενο πλήθος ψηφίων της εξόδου $\mathbb{E}[|Com^*(\sigma)|] \leq (1+\delta)nH(p)$.

(β) Για κάθε συνάρτηση συμπίεσης Com, το αναμενόμενο πλήθος ψηφίων της εξόδου είναι $\mathbb{E}[|Com(\sigma)|] \geq (1-\delta)nH(p)$.

Συσχέτιση Εντροπίας και Συμπίεσης (I)

ΘΕΩΡΗΜΑ 9.6 [Mitzenmacher-Upfal (2005), Chapter 9]

Έστω «πειραγμένο» νόμισμα $COIN(p)$ με $\mathbb{P}[H] = p > 1/2$. $\forall \delta > 0$ και αρκούντως μεγάλο $n \in \mathbb{N}$, ως προς το αποτέλεσμα $\sigma \in \{H,T\}^n$ των n ανεξάρτητων ρίψεων του $COIN(p)$:

(α) Υπάρχει συνάρτηση συμπίεσης Com^* τ.ώ. με αναμενόμενο πλήθος ψηφίων της εξόδου $\mathbb{E}[|Com^*(\sigma)|] \leq (1+\delta)nH(p)$.

(β) Για κάθε συνάρτηση συμπίεσης Com, το αναμενόμενο πλήθος ψηφίων της εξόδου είναι $\mathbb{E}[|Com(\sigma)|] \geq (1-\delta)nH(p)$.

Π1 Αν θεωρούσαμε αυθαιρετους **κώδικες προθέματος** Com, το κάτω φράγμα θα μπορούσε να γίνει $\mathbb{E}[|Com(\sigma)|] \geq nH(p)$.

Π2 Η συνάρτηση συμπίεσης Com^* που χρησιμοποιείται στην απόδειξη του άνω φράγματος δημιουργεί έναν **κώδικα προθέματος**.

Συσχέτιση Εντροπίας και Συμπίεσης (II)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΑΝΩ ΦΡΑΓΜΑ

- **Βασική Ιδέα:** Το πλήθος των H στο σ είναι συγκεντρωμένο γύρω από τη μέση τιμή np.
- $\sigma \in \{H,T\}^n$: Η είσοδος (μεταβλητή).

Συσχέτιση Εντροπίας και Συμπίεσης (II)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΑΝΩ ΦΡΑΓΜΑ

- **Βασική Ιδέα:** Το πλήθος των H στο σ είναι συγκεντρωμένο γύρω από τη μέση τιμή np.
- $\sigma \in \{H,T\}^n$: Η είσοδος (μεταβλητή).
- $Com^*(\sigma) =$

$$\begin{cases} 0\tau, & \sum_{k=1}^{n} \mathbb{I}_{[\sigma[k]=H]} \geq n(p-\epsilon), & \tau \in \{0,1\}^{nH(p-\epsilon)+\log_2(n)} \\ 1\tau, & \sum_{k=1}^{n} \mathbb{I}_{[\sigma[k]=H]} \leq n(p-\epsilon)-1, & \tau \in \{0,1\}^{n} \text{ /* καθόλου συμπίεση */} \end{cases}$$

Συσχέτιση Εντροπίας και Συμπίεσης (II)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΑΝΩ ΦΡΑΓΜΑ

- **Βασική Ιδέα:** Το πλήθος των H στο σ είναι συγκεντρωμένο γύρω από τη μέση τιμή np.
- $\sigma \in \{H,T\}^n$: Η είσοδος (μεταβλητή).
- $Com^*(\sigma) =$
$$\begin{cases} 0\tau, & \sum_{k=1}^{n} \mathbb{I}_{[\sigma[k]=H]} \geq n(p-\epsilon), & \tau \in \{0,1\}^{nH(p-\epsilon)+\log_2(n)} \\ 1\tau, & \sum_{k=1}^{n} \mathbb{I}_{[\sigma[k]=H]} \leq n(p-\epsilon)-1, & \tau \in \{0,1\}^{n} \end{cases}$$ /* καθόλου συμπίεση */
- $\underline{Com^*(\sigma)[1]=1.}$ $\mathbb{P}[Com^*(\sigma)[1]=1] \leq \exp\left(-\frac{pn\epsilon^2}{2}\right)$ /* Chernoff Bound */

Συσχέτιση Εντροπίας και Συμπίεσης (II)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΑΝΩ ΦΡΑΓΜΑ

- **Βασική Ιδέα:** Το πλήθος των H στο σ είναι συγκεντρωμένο γύρω από τη μέση τιμή np.
- $\sigma \in \{H,T\}^n$: Η είσοδος (μεταβλητή).
- $Com^*(\sigma) =$
$$\begin{cases} 0\tau, & \sum_{k=1}^{n} \mathbb{I}_{[\sigma[k]=H]} \geq n(p-\epsilon), & \tau \in \{0,1\}^{nH(p-\epsilon)+\log_2(n)} \\ 1\tau, & \sum_{k=1}^{n} \mathbb{I}_{[\sigma[k]=H]} \leq n(p-\epsilon)-1, & \tau \in \{0,1\}^{n} \end{cases}$$ /* καθόλου συμπίεση */
- $\underline{Com^*(\sigma)[1]=1.}$ $\mathbb{P}[Com^*(\sigma)[1]=1] \leq \exp\left(-\frac{pn\epsilon^2}{2}\right)$ /* Chernoff Bound */
- $\underline{Com^*(\sigma)[1]=0.}$ Για $p-\epsilon > \frac{1}{2} \Leftrightarrow n(p-\epsilon) > \frac{n}{2}$,
$|\{\sigma : \sigma[1]=1\}| = \sum_{j=\lceil n(p-\epsilon)\rceil}^{n} C(n,j) \leq \frac{n}{2} C(n, \lceil n(p-\epsilon)\rceil)$
$\leq 2^{nH(p-\epsilon)+\log_2(n)}$

Συσχέτιση Εντροπίας και Συμπίεσης (II)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΑΝΩ ΦΡΑΓΜΑ

- **Βασική Ιδέα:** Το πλήθος των H στο σ είναι συγκεντρωμένο γύρω από τη μέση τιμή np.
- $\sigma \in \{H,T\}^n$: Η είσοδος (μεταβλητή).
- $Com^*(\sigma) =$
$$\begin{cases} 0\tau, & \sum_{k=1}^{n} \mathbb{I}_{[\sigma[k]=H]} \geq n(p-\epsilon), & \tau \in \{0,1\}^{nH(p-\epsilon)+\log_2(n)} \\ 1\tau, & \sum_{k=1}^{n} \mathbb{I}_{[\sigma[k]=H]} \leq n(p-\epsilon)-1, & \tau \in \{0,1\}^{n} \end{cases}$$ /* καθόλου συμπίεση */
- $\underline{Com^*(\sigma)[1]=1.}$ $\mathbb{P}[Com^*(\sigma)[1]=1] \leq \exp\left(-\frac{pn\epsilon^2}{2}\right)$ /* Chernoff Bound */
- $\underline{Com^*(\sigma)[1]=0.}$ Για $p-\epsilon > \frac{1}{2} \Leftrightarrow n(p-\epsilon) > \frac{n}{2}$,
$|\{\sigma : \sigma[1]=1\}| = \sum_{j=\lceil n(p-\epsilon)\rceil}^{n} C(n,j) \leq \frac{n}{2} C(n, \lceil n(p-\epsilon)\rceil)$
$\leq 2^{nH(p-\epsilon)+\log_2(n)}$
- $\mathbb{E}[Com^*(\Sigma)] \leq (n+1)\cdot \exp\left(-\frac{pn\epsilon^2}{2}\right) + (nH(p-\epsilon)+\log_2(n)+1)$
$\leq (1+\delta)nH(p)$ /* για καταλληλη επιλογή $\epsilon > 0$ και $n \in \mathbb{N}$ */

Συσχέτιση Εντροπίας και Συμπίεσης (II)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΚΑΤΩ ΦΡΑΓΜΑ

- $\forall \Sigma_1, \Sigma_2 \in \{H,T\}^n,\ \#H(\Sigma_1) > \#H(\Sigma_2) \ \rightarrow\ \mathbb{P}[\sigma = \Sigma_1] > \mathbb{P}[\sigma = \Sigma_2]$.

Συσχέτιση Εντροπίας και Συμπίεσης (II)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΚΑΤΩ ΦΡΑΓΜΑ

- $\forall \Sigma_1, \Sigma_2 \in \{H,T\}^n,\ \#H(\Sigma_1) > \#H(\Sigma_2) \ \rightarrow\ \mathbb{P}[\sigma = \Sigma_1] > \mathbb{P}[\sigma = \Sigma_2]$.
- L9.7 (Mitzenmacher-Upfal): Για συνάρτηση συμπίεσης *Com* που **ελαχιστοποιεί** το αναμενόμενο μήκος αναπαράστασης, $\mathbb{P}[\sigma = \Sigma_1] > \mathbb{P}[\sigma = \Sigma_2] \ \rightarrow\ |Com(\Sigma_2)| \geq |Com(\Sigma_1)|$.

Συσχέτιση Εντροπίας και Συμπίεσης (II)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΚΑΤΩ ΦΡΑΓΜΑ

- $\forall \Sigma_1, \Sigma_2 \in \{H,T\}^n,\ \#H(\Sigma_1) > \#H(\Sigma_2) \ \rightarrow\ \mathbb{P}[\sigma = \Sigma_1] > \mathbb{P}[\sigma = \Sigma_2]$.
- L9.7 (Mitzenmacher-Upfal): Για συνάρτηση συμπίεσης *Com* που **ελαχιστοποιεί** το αναμενόμενο μήκος αναπαράστασης, $\mathbb{P}[\sigma = \Sigma_1] > \mathbb{P}[\sigma = \Sigma_2] \ \rightarrow\ |Com(\Sigma_2)| \geq |Com(\Sigma_1)|$.
- Αν η *Com* χρησιμοποιεί διαφορετικές δυαδικές συμβολοσειρές για s εισόδους από το $\{H,T\}^n$, τότε τουλάχιστον μια έξοδος έχει μήκος $\geq \lceil \log_2(s+1) \rceil - 1$.

Συσχέτιση Εντροπίας και Συμπίεσης (II)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΚΑΤΩ ΦΡΑΓΜΑ

- $\forall \Sigma_1, \Sigma_2 \in \{H,T\}^n,\ \#H(\Sigma_1) > \#H(\Sigma_2) \ \rightarrow\ \mathbb{P}[\sigma = \Sigma_1] > \mathbb{P}[\sigma = \Sigma_2]$.
- L9.7 (Mitzenmacher-Upfal): Για συνάρτηση συμπίεσης *Com* που **ελαχιστοποιεί** το αναμενόμενο μήκος αναπαράστασης, $\mathbb{P}[\sigma = \Sigma_1] > \mathbb{P}[\sigma = \Sigma_2] \ \rightarrow\ |Com(\Sigma_2)| \geq |Com(\Sigma_1)|$.
- Αν η *Com* χρησιμοποιεί διαφορετικές δυαδικές συμβολοσειρές για s εισόδους από το $\{H,T\}^n$, τότε τουλάχιστον μια έξοδος έχει μήκος $\geq \lceil \log_2(s+1) \rceil - 1$.
- Από L9.3: Το πλήθος εισόδων με $s = \lfloor n(p+\epsilon) \rfloor$ εμφανίσεις H είναι $C(n, \lfloor n(p+\epsilon) \rfloor) \geq \frac{2^{nH(p+\epsilon)}}{n+1} = 2^{nH(p+\epsilon) - \log_2(n+1)}$.

Συσχέτιση Εντροπίας και Συμπίεσης (III)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΚΑΤΩ ΦΡΑΓΜΑ

- $A = \log_2 \left(2^{nH(p+\epsilon)-\log_2(n+1)} + 1 \right) - 1 \geq nH(p+\epsilon) - \log_2(n+1)$ ψηφία, για τουλάχιστον μια είσοδο με $n(p+\epsilon)$ εμφανίσεις του Η.

Συσχέτιση Εντροπίας και Συμπίεσης (III)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΚΑΤΩ ΦΡΑΓΜΑ

- $A = \log_2\left(2^{nH(p+\epsilon)-\log_2(n+1)}+1\right)-1 \geq nH(p+\epsilon)-\log_2(n+1)$ ψηφία, για τουλάχιστον μια είσοδο με $n(p+\epsilon)$ εμφανίσεις του H.
- Τουλάχιστον A ψηφία, για κάθε είσοδο με λιγότερες από $n(p+\epsilon)$ εμφανίσεις του H.

Συσχέτιση Εντροπίας και Συμπίεσης (III)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΚΑΤΩ ΦΡΑΓΜΑ

- $A = \log_2\left(2^{nH(p+\epsilon)-\log_2(n+1)}+1\right)-1 \geq nH(p+\epsilon)-\log_2(n+1)$ ψηφία, για τουλάχιστον μια είσοδο με $n(p+\epsilon)$ εμφανίσεις του H.
- Τουλάχιστον A ψηφία, για κάθε είσοδο με λιγότερες από $n(p+\epsilon)$ εμφανίσεις του H.
- Chernoff Bound:
  $\mathbb{P}$ [τουλάχιστον $n(p+\epsilon)$ εμφανίσεις του H] $< exp\left(-\frac{pn\epsilon^2}{12}\right)$

Συσχέτιση Εντροπίας και Συμπίεσης (III)

ΕΞΗΓΗΣΗ: (ΘΕΩΡΗΜΑ 9.6) ΚΑΤΩ ΦΡΑΓΜΑ

- $A = \log_2\left(2^{nH(p+\epsilon)-\log_2(n+1)}+1\right)-1 \geq nH(p+\epsilon)-\log_2(n+1)$ ψηφία, για τουλάχιστον μια είσοδο με $n(p+\epsilon)$ εμφανίσεις του H.
- Τουλάχιστον A ψηφία, για κάθε είσοδο με λιγότερες από $n(p+\epsilon)$ εμφανίσεις του H.
- Chernoff Bound:
  $\mathbb{P}$ [τουλάχιστον $n(p+\epsilon)$ εμφανίσεις του H] $< exp\left(-\frac{pn\epsilon^2}{12}\right)$
- Εφόσον η Com ελαχιστοποιεί το αναμενόμενο μήκος κωδικοποίησης, ισχύει το εξής:
  $\mathbb{E}[|Com(\sigma)|] \geq [nH(p+\epsilon)-\log_2(n+1)]\cdot\left[1-exp\left(-\frac{pn\epsilon^2}{12}\right)\right]$
  $\geq (1-\delta)nH(p)$

Σκελετός Ομιλίας

Θεωρία Κωδίκων

- **Συμπίεση:** Μειώνει το αναμενόμενο πλήθος ψηφίων για την αναπαράσταση πληροφορίας.

Θεωρία Κωδίκων

- **Συμπίεση:** Μειώνει το αναμενόμενο πλήθος ψηφίων για την αναπαράσταση πληροφορίας.
- **Κωδικοποίηση:** Εισάγει επιπρόσθετη πληροφορία για την αντιμετώπιση σφαλμάτων,

Θεωρία Κωδίκων

- **Συμπίεση:** Μειώνει το αναμενόμενο πλήθος ψηφίων για την αναπαράσταση πληροφορίας.
- **Κωδικοποίηση:** Εισάγει επιπρόσθετη πληροφορία για την αντιμετώπιση σφαλμάτων,
  - Η πληροφορία μεταδίδεται από έναν αποστολέα προς έναν παραλήπτη, μέσω ενός «θορυβώδους» καναλιού.

Θεωρία Κωδίκων

- **Συμπίεση:** Μειώνει το αναμενόμενο πλήθος ψηφίων για την αναπαράσταση πληροφορίας.
- **Κωδικοποίηση:** Εισάγει επιπρόσθετη πληροφορία για την αντιμετώπιση σφαλμάτων,
  - Η πληροφορία μεταδίδεται από έναν αποστολέα προς έναν παραλήπτη, μέσω ενός «θορυβώδους» καναλιού.
  - Συμβαίνει διαταραχή κάποιων ψηφίων.

Θεωρία Κωδίκων

- **Συμπίεση:** Μειώνει το αναμενόμενο πλήθος ψηφίων για την αναπαράσταση πληροφορίας.
- **Κωδικοποίηση:** Εισάγει επιπρόσθετη πληροφορία για την αντιμετώπιση σφαλμάτων,
  - Η πληροφορία μεταδίδεται από έναν αποστολέα προς έναν παραλήπτη, μέσω ενός «θορυβώδους» καναλιού.
  - Συμβαίνει διαταραχή κάποιων ψηφίων.
  - ΣΤΟΧΟΣ: Ανάκτηση μεταδοθείσας πληροφορίας, παρά τον θόρυβο.

ΟΡΙΣΜΟΣ 9.4

[Mitzenmacher-Upfal (2005), Chapter 9]

Δυαδικό συμμετρικό κανάλι: Δέχεται ως είσοδο μια συμβολοσειρά $x \in \{0,1\}^m$ και επιστρέφει ως έξοδο μια συμβολοσειρά $y \in \{0,1\}^m$ τ.ώ. $\forall k \in [m],\ \mathbb{P}[y_k = x_k] = 1 - p$, ανεξάρτητα για κάθε ψηφίο.

ΟΡΙΣΜΟΣ 9.4

[Mitzenmacher-Upfal (2005), Chapter 9]

Δυαδικό συμμετρικό κανάλι: Δέχεται ως είσοδο μια συμβολοσειρά $x \in \{0,1\}^m$ και επιστρέφει ως έξοδο μια συμβολοσειρά $y \in \{0,1\}^m$ τ.ώ. $\forall k \in [m],\ \mathbb{P}[y_k = x_k] = 1 - p$, ανεξάρτητα για κάθε ψηφίο.

Q8 Πώς θα μπορούσαμε να προστατέψουμε τη μετάδοση ενός ψηφίου $x \in \{0,1\}$ πληροφορίας, από το συμμετρικό κανάλι $CHANNEL(p)$ για $p < 1/2$;

ΟΡΙΣΜΟΣ 9.4

[Mitzenmacher-Upfal (2005), Chapter 9]

Δυαδικό συμμετρικό κανάλι: Δέχεται ως είσοδο μια συμβολοσειρά $x \in \{0,1\}^m$ και επιστρέφει ως έξοδο μια συμβολοσειρά $y \in \{0,1\}^m$ τ.ώ. $\forall k \in [m],\ \mathbb{P}[y_k = x_k] = 1 - p$, ανεξάρτητα για κάθε ψηφίο.

Q8 Πώς θα μπορούσαμε να προστατέψουμε τη μετάδοση ενός ψηφίου $x \in \{0,1\}$ πληροφορίας, από το συμμετρικό κανάλι $CHANNEL(p)$ για $p < 1/2$;

A8 Στέλνουμε n φορές το x από το κανάλι. Επιστρέφουμε την τιμή που εμφανίστηκε περισσότερες φορές. Η πιθανότητα εσφαλμένης απάντησης, καθώς $n \to \infty$, τείνει στο μηδέν.

Συναρτήσεις Κωδικοποίησης

ΟΡΙΣΜΟΣ 9.5

[Mitzenmacher-Upfal (2005), Chapter 9]

- (k,n)-**Συνάρτηση κωδικοποίησης** $Enc : \{0,1\}^k \mapsto \{0,1\}^n$: Δέχεται ως είσοδο μια k-ψήφια δυαδική συμβολοσειρά, και παράγει στην έξοδο μια n-ψήφια δυαδική συμβολοσειρά.
- (k,n)-**Συνάρτηση αποκωδικοποίησης** $Dec : \{0,1\}^n \mapsto \{0,1\}^k$: Δέχεται ως είσοδο μια n-ψήφια δυαδική συμβολοσειρά, και παράγει στην έξοδο μια k-ψήφια δυαδική συμβολοσειρά.
- (k,n)-**Σχήμα κωδικοποίησης**: $Enc, Dec : \{0,1\}^k \mapsto \{0,1\}^n$ τ.ώ. $\forall \sigma \in \{0,1\}^k, Dec(Enc(\sigma)) = \sigma$.

Συναρτήσεις Κωδικοποίησης

ΟΡΙΣΜΟΣ 9.5

[Mitzenmacher-Upfal (2005), Chapter 9]

- (k,n)-**Συνάρτηση κωδικοποίησης** $Enc : \{0,1\}^k \mapsto \{0,1\}^n$: Δέχεται ως είσοδο μια k-ψήφια δυαδική συμβολοσειρά, και παράγει στην έξοδο μια n-ψήφια δυαδική συμβολοσειρά.
- (k,n)-**Συνάρτηση αποκωδικοποίησης** $Dec : \{0,1\}^n \mapsto \{0,1\}^k$: Δέχεται ως είσοδο μια n-ψήφια δυαδική συμβολοσειρά, και παράγει στην έξοδο μια k-ψήφια δυαδική συμβολοσειρά.
- (k,n)-**Σχήμα κωδικοποίησης**: $Enc, Dec : \{0,1\}^k \mapsto \{0,1\}^n$ τ.ώ. $\forall \sigma \in \{0,1\}^k, Dec(Enc(\sigma)) = \sigma$.

- ΣΤΟΧΟΣ: Εύρεση Σχήματος Κωδικοποίησης (Enc, Dec) τ.ώ.

$$\forall \sigma \in \{0,1\}^k, \ \mathbb{P}\left[Dec(CHANNEL(p, Enc(\sigma))) \neq \sigma\right] \in o(1)$$

Θεώρημα του Shanon

ΘΕΩΡΗΜΑ 9.8 [Mitzenmacher-Upfal (2005), Chapter 9]

Για ένα δυαδικό συμμετρικό κανάλι $CHANNEL(p)$ με $p \in (0, 1/2)$, οποιεσδήποτε σταθερές $\gamma, \delta > 0$, και αρκούντως μεγάλο $n \in \mathbb{N}$ ισχύουν τα εξής:

(α) Για κάθε $k \leq n(1 - H(p))$, υπάρχουν συναρτήσεις (k, n)-κωδικοποίησης / αποκωδικοποίησης τ.ώ. η πιθανότητα παραλαβής εσφαλμένου μηνύματος είναι το πολύ γ, για οποιαδηποτε είσοδο $x \in \{0, 1\}^k$.

(β) Δεν υπάρχουν συναρτήσεις (k, n)-κωδικοποίησης / αποκωδικοποίησης με $k \geq n(1 - H(p) + \delta)$ τ.ώ. η πιθανότητα ορθής αποκωδικοποίησης να είναι τουλάχιστον γ, για είσοδο $x \in_{\mathbf{uar}} \{0, 1\}^k$.

ΠΑΡΑΔΕΙΓΜΑ

1η Ανάθεση Για Θ. Πληροφορίας (2015) (I)

1. Για κάθε ζεύγος $n-$ψήφιων συμβολοσειρών $\mathbf{y}, \mathbf{z} \in \{0,1\}^n$ ορίζουμε την **απόσταση Hamming** ως εξής: $\Delta(\mathbf{y}, \mathbf{z}) = \sum_{j=1}^{n} |y_j - z_j|$.

2. Έστω ότι επιθυμούμε να μεταδωσουμε $n-$ψήφιες δυαδικές συμβολοσειρές $\mathbf{z} \in \{0,1\}^n$ μέσω ενό **δυαδικού συμμετρικού καναλιού** $\mathbf{y} = CHANNEL_p(\mathbf{z}) \in \{0,1\}^n$ με πιθανότητα αναστροφής κάθε ψηφίου $p \in (0, 1/2)$.

3. Ακολουθούμε ένα σχήμα *κωδικοποίησης − μετάδοσης − αποκωδικοποίησης* (βλ. επόμενη διαφάνεια) προκειμένου να μεταδώσουμε $k-$ψήφιες δυαδικές συμβολοσειρές με τέτοιο τρόπο ώστε να ελαχιστοποιείται η πιθανότητα αποτυχίας αποκωδικοποίησης από την πλευρά του παραλήπτη.

ΠΑΡΑΔΕΙΓΜΑ

1η Ανάθεση Για Θ. Πληροφορίας (2015) (II)

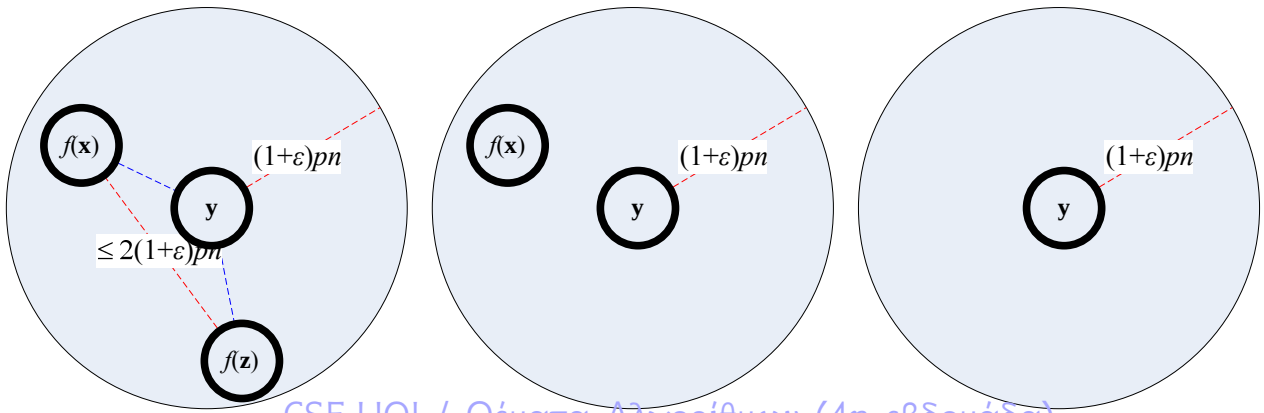
4. Σχήμα κωδικοποίησης / αποκωδικοποίησης:

Κωδικοποίηση: Δημιουργείται μια *ομοιόμορφη αντιστοιχία* $k-$ψήφιων δυαδικών συμβολοσειρών σε $n-$ψήφιες δυαδικές συμβολοσειρές ($n \geq k \geq 1$): $\forall \mathbf{x} \in \{0,1\}^k,\ f(\mathbf{x}) \in_{\mathbf{uar}} \{0,1\}^n$

Μετάδοση: Για αυθαίρετο μήνυμα $\mathbf{x} \in \{0,1\}^k$, μεταδίδεται το $f(\mathbf{x})$ μέσω του δυαδικού συμμετρικού καναλιού: $\mathbf{y} = CHANNEL_p(f(\mathbf{x})) \in \{0,1\}^n$

Αποκωδικοποίηση: $N(\mathbf{y}) = \left\{\mathbf{x}' \in \{0,1\}^k : \Delta(\mathbf{y}, f(\mathbf{x}')) \leq (1+\epsilon)pn\right\}$

if $|N(\mathbf{y})| = 1$ **then return** $(N(\mathbf{y}))$ **else return** «αποτυχία»



ΠΑΡΑΔΕΙΓΜΑ

1η Ανάθεση Για Θ. Πληροφορίας (2015) (III)

ΘΠ.1 Για αυθαίρετα $\mathbf{x}, \mathbf{x}' \in \{0,1\}^k : \mathbf{x} \neq \mathbf{x}'$, να υπολογιστεί η αναμενόμενη τιμή $\mathbb{E}[\Delta(f(\mathbf{x}),f(\mathbf{x}'))]$, αλλά και η πιθανότητα $\mathbb{P}\left[\Delta(f(\mathbf{x}),f(\mathbf{x}')) < (1-\delta)\frac{n}{2}\right]$, για κάποιο $\delta \in (0,1)$.

ΘΠ.2 Για αυθαίρετο $\mathbf{z} \in \{0,1\}^n$, να υπολογιστεί η αναμενόμενη τιμή $\mathbb{E}[\Delta(\mathbf{z},\mathbf{y})]$ όπου $\mathbf{y} = CHANNEL_p(\mathbf{z})$.

ΘΠ.3 Για αυθαίρετο $\mathbf{x} \in \{0,1\}^k$ και $\mathbf{y} = CHANNEL_p(f(\mathbf{x}))$, νδο $\mathbb{P}[f(\mathbf{x}) \notin N(\mathbf{y})] \leq \exp\left(-\frac{pn\epsilon^2}{3}\right)$.

ΘΠ.4 Για αυθαίρετο $\mathbf{x} \in \{0,1\}^k$, νδο $\mathbb{P}[\exists \mathbf{x}' \neq \mathbf{x} : \Delta(f(\mathbf{x}),f(\mathbf{x}')) < 2(1+\epsilon)pn] \leq \exp\left(k\ln(2) - \frac{[1-4(1+\epsilon)p]^2 n}{6}\right)$

ΘΠ.5 Καταλήξτε ότι αρκεί να θέσουμε $n = \frac{6\cdot[k\ln(2)+\ln(k)]}{[1-4(1+\epsilon)p]^2}$ προκειμένου να εξασφαλίσουμε πιθανότητα αποτυχίας της αποκωδικοποίησης $O\left(\frac{1}{k}\right)$.

ΠΑΡΑΔΕΙΓΜΑ

1η Ανάθεση Για Θ. Πληροφορίας (2015) (IV)

Βασικά Εργαλεία Από Θεωρία Πιθανοτήτων:

- **UNION BOUND (UB):** Για κάθε πεπερασμένο (ή αριθμήσιμα άπειρο) συνολο γεγονότων $A_1, A_2, \ldots$, η πιθανότητα να συμβεί *τουλάχιστον ένα από αυτά τα γεγονότα* φράσσεται άνω από το άθροισμα των πιθανοτήτων των γεγονότων αυτών:

$$\mathbb{P}\left[\left(\bigcup_i A_i\right)\right] \leq \sum_i \mathbb{P}[A_i]$$

- **CHERNOFF BOUND (CB):** Για ανεξάρτητες, ομοιόμορφα κατανεμημένες $\{0,1\}-$μεταβλητές $X_1, X_2, \ldots, X_m$ τ.ώ. $\mathbb{E}[X_i] = \mu$, $X = \sum_{i=1}^{m} X_i$ και $\delta \in (0,1)$, ισχύουν τα εξής:

$$\textbf{CB1} \quad \mathbb{P}[X \geq (1+\delta)\mu m] \leq \exp\left(-\frac{\mu m \delta^2}{3}\right)$$

$$\textbf{CB2} \quad \mathbb{P}[X \leq (1-\delta)\mu m] \leq \exp\left(-\frac{\mu m \delta^2}{2}\right)$$

Ευχαριστώ για την προσοχή σας!

Ερωτήσεις / Σχόλια ;