Λυμένες Ασκήσεις σε Εντροπία, Αμοιβαία Πληροφορία, Κωδικοποίηση Πηγής και AEP

1. **► Συναρτήσεις τυχαίων μεταβλητών – Cover & Thomas 2.2 (παραλλαγή)**

   (*α) Έστω $y = g(x)$, $g(\cdot)$ ντετερμινιστική συνάρτηση και X μια (διακριτή) τυχαία μεταβλητή. Μπορεί να ειπωθεί κάτι για τη σχέση μεταξύ της εντροπίας της $Y = g(X)$ και της εντροπίας της X; Εάν ναι, τι;

   Υπόδειξη: Ο πιο απλός τρόπος να λύσετε την άσκηση είναι με χρήση της αρχής διαχωρισιμότητας της εντροπίας.

   (β) Βρείτε τη σχέση μεταξύ $H(X)$ και $H(Y)$ για τις συναρτήσεις $y = x^3$ και $y = \lfloor x/2 \rfloor$, x ακέραιος.

2. **► Εντροπία αθροίσματος τυχαίων μεταβλητών – Cover & Thomas 2.14**

   (α) Έστω X και Y τυχαίες μεταβλητές και $Z = X + Y$. Να αποδειχτεί ότι $H(Z|X) = H(Y|X)$.

   (*β) Εάν οι X και Y είναι *ανεξάρτητες*, να αποδειχτεί ότι $H(Y) \leq H(Z)$ και $H(X) \leq H(Z)$. Επομένως, όταν σε μια τ.μ. προστίθεται μια ανεξάρτητή της τ.μ., η αβεβαιότητα αυξάνει.

   (*γ) Δώστε ένα παράδειγμα (μη ανεξάρτητων) τ.μ. X και Y για τις οποίες $H(Y) \geq H(Z)$ και $H(X) \geq H(Z)$.

3. **Μια εναλλακτική απόδειξη της ανισότητας $D(p||q) \geq 0$ – Cover & Thomas 2.26**

   (α) Δείξτε ότι, για $0 < x < \infty$, $\ln x \leq x - 1$.

   (β) Δικαιολογήστε τα παρακάτω βήματα

$$\begin{aligned} -D(p||q) &= \sum_x p(x) \ln \frac{q(x)}{p(x)} \\ &\overset{(i)}{\leq} \sum_x p(x) \left(\frac{q(x)}{p(x)} - 1 \right) \\ &\overset{(ii)}{\leq} 0. \end{aligned}$$

   (γ) Πότε ισχύει η ισότητα;

4. **Ανισότητα Fano – Cover & Thomas 2.32**

Έστω η εξής από κοινού κατανομή μάζας πιθανότητας για τις τ.μ. (X, Y)

		Y		
		a	b	c
X	1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
	2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
	3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Έστω, επίσης, $\hat{X}(Y)$ εκτιμητής για τη X με βάση την Y και $P_e = \Pr\{\hat{X}(Y) \neq X\}$.

(α) Βρείτε το βέλτιστο εκτιμητή $\hat{X}(Y)$, καθώς και την P_e που αντιστοιχεί στο βέλτιστο εκτιμητή.

(β) Βρείτε το φράγμα για την P_e που προκύπτει από την ανισότητα Fano και συγκρίνετε.

5. **Σχετική Εντροπία – Cover & Thomas 2.37**

Έστω 3 τ.μ. X, Y, Z με από κοινού συνάρτηση μάζας πιθανότητας $p(x, y, z)$. Η σχετική εντροπία μεταξύ της από κοινού κατανομής και του γινομένου των περιθώριων κατανομών ορίζεται ως

$$D(p(x,y,z)||p(x)p(y)p(z)) = \mathbb{E}\left[\log \frac{p(x,y,z)}{p(x)p(y)p(z)}\right].$$

(α) Εκφράστε την $D(p(x,y,z)||p(x)p(y)p(z))$ συναρτήσει εντροπιών.

(β) Πότε η $D(p(x,y,z)||p(x)p(y)p(z))$ ισούται με 0;

6. **Κέρμα και ζάρι – Cover & Thomas 2.43**

(α) Θεωρήστε το αποτέλεσμα της ρίψης *αμερόληπτου* κέρματος. Με τι ισούται η αμοιβαία πληροφορία μεταξύ της επάνω και κάτω πλευράς του κέρματος; Συμφωνεί το αποτέλεσμα με αυτό που περιμένατε διαισθητικά;

(β) Αλλάζει η απάντησή σας εάν το κέρμα είναι μεροληπτικό;

(γ) Θεωρήστε τη ρίψη αμερόληπτου ζαριού με 6 πλευρές. Ποια είναι η αμοιβαία πληροφορία μεταξύ της επάνω και της μπροστινής πλευράς; Υπενθυμίζεται ότι σε ένα ζάρι το άθροισμα των αντίθετων πλευρών ισούται με 7.

7. **Μήκος Ακολουθίας – Cover & Thomas 2.48**

Θεωρούμε τυχαία διαδικασία Bernoulli$\left(\frac{1}{2}\right)$ $\{X_i\}$. Σταματάμε τη διαδικασία όταν εμφανίζεται το πρώτο "1". Έστω N το μήκος της ακολουθίας όταν σταματάμε. Επομένως, η ακολουθία X^N είναι ένα στοιχείο του συνόλου όλων των δυαδικών ακολουθιών πεπερασμένου μήκους: $\{0,1\}^* = \{0, 1, 00, 01, 10, 11, 000, \ldots\}$.

(α) Βρείτε την $I(N; X^N)$.

(β) Βρείτε την $H(X^N|N)$.

(γ) Βρείτε την $H(X^N)$.

(δ) Βρείτε την $H(N|X^N)$.

(ε) Βρείτε την $H(N)$.

Υπόδειξη: Ενδέχεται να σας φανεί χρήσιμη η Άσκηση 2.1 του βιβλίου των Cover & Thomas.

Αλλάζουμε, τώρα, τον τρόπο με τον οποίο σταματάμε την παραγωγή της ακολουθίας. Θεωρήστε και πάλι ότι $X_i \sim$Bernoulli$\left(\frac{1}{2}\right)$. Με πιθανότητα $\frac{1}{3}$ σταματάμε όταν $N = 6$, αλλιώς σταματάμε όταν $N = 12$. Επίσης, θεωρούμε ότι η επιλογή της τιμής του N είναι *ανεξάρτητη* της ακολουθίας $X_1X_2 \ldots X_{12}$.

(στ) Βρείτε την $I(N; X^N)$.

(ζ) Βρείτε την $H(X^N|N)$.

(η) Βρείτε την $H(X^N)$.

(θ) Βρείτε την $H(N|X^N)$.

(ι) Βρείτε την $H(N)$.

8. **Εντροπία και Αμοιβαία Πληροφορία (Πρόοδος Θ. Π., Νοέμβριος 2007)**

Έστω οι *ανεξάρτητες* τ.μ. X και Y. Η X ακολουθεί ομοιόμορφη κατανομή με τιμές $\{-1, 1\}$ (δηλαδή $X = 1$ ή $X = -1$ με την ίδια πιθανότητα), ενώ η Y ακολουθεί ομοιόμορφη κατανομή με *γνωστές* τιμές $\{-a, a\}$, $a \in \mathbb{R}$ (δηλαδή $Y = a$ ή $Y = -a$ με την ίδια πιθανότητα). Η τ.μ. Z δίνεται από τη σχέση $Z = X + Y$. Για όλες τις πιθανές τιμές του $a \in \mathbb{R}$ να βρεθούν

(α) Η $H(Y)$.

(β) Η $H(Z)$.

(γ) Η $I(X; Z)$.

9. **Εντροπία, Αμοιβαία Πληροφορία και συμπίεση (Τελικό διαγώνισμα Θ. Π., Φεβρουάριος 2008)**

Έστω οι *ανεξάρτητες* τ.μ. X και Y, η καθεμία με ομοιόμορφη κατανομή στο σύνολο $\{0, 1, 2, 3\}$. Η τ.μ. Z δίνεται από τη σχέση $Z = X \cdot Y$ (γινόμενο). Να βρεθούν

(α) Η $H(X)$.

(β) Η $H(Z)$. $\log_2 7 \approx 2.8074$.

(γ) Η $I(X; Z)$.

(δ) Κατασκευάστε ένα δυαδικό κώδικα Huffman για την τ.μ. Z. Συγκρίνετε το μέσο μήκος κώδικα με την εντροπία της Z και σχολιάστε.

10. **Σύντομες Ερωτήσεις (Τελικό Διαγώνισμα Θ. Π., Φεβρουάριος 2008)**

(α) Αποδείξτε ότι $H(X, Y, Z) - H(X, Y) \leq H(X, Z) - H(X)$. Πότε ισχύει η ισότητα;

(β) Σωστό ή λάθος; $I(X; Y|Z) = 0 \Rightarrow I(X; Y) = 0$. Αποδείξτε ή δώστε αντιπαράδειγμα.

11. **Εντροπία (Τελικό Διαγώνισμα Π. Θ. Θ. Π., Ιούνιος 2008)**

    Θεωρούμε μια διακριτή τυχαία τ.μ. που παίρνει τιμές από ένα πεπερασμένο σύνολο στοιχείων $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$. Έστω ότι δίνεται η $p(1) = \Pr\{X = 1\}$ η οποία δεν είναι δυνατόν να μεταβληθεί.

    (α) Ποιες είναι οι τιμές $p(i)$, $i = 2, \ldots, |\mathcal{X}|$, της κατανομής που μεγιστοποιεί την εντροπία $H(X)$ (για δεδομένη $p(1)$);

    (β) Με τι ισούται η μέγιστη $H(X)$ για δεδομένη $p(1)$; Δώστε μια ερμηνεία της έκφρασης για την $H(X)$ με χρήση της αρχής διαχωρισιμότητας.

    Υπόδειξη: Χωρίστε το $\mathcal{X}$ σε δύο κατάλληλα υποσύνολα.

    (γ) Εάν ένας παίκτης Α μπορεί να μεταβάλλει την $p(1)$ με σκοπό να μειώνει την εντροπία $H(X)$ όσο περισσότερο μπορεί, ενώ ένας παίκτης Β μπορεί να μεταβάλλει οποιοδήποτε υποσύνολο των υπόλοιπων $p(i)$ (αλλά όχι την $p(1)$) με σκοπό να αυξάνει την εντροπία όσο μπορεί, ποια τιμή πρέπει να επιλέξει ο Α για την $p(1)$ αν παίζει πρώτος; Πώς πρέπει να απαντήσει ο Β; Αλλάζει η απάντησή σας εάν πρώτος παίζει ο Β; Θεωρούμε ότι οι παίκτες παίζουν έτσι ώστε να μην παραβιάζεται η συνθήκη $\sum_{i=1}^{|\mathcal{X}|} p(x_i) = 1$ για το **p**.

12. **Εντροπία (Επαναληπτική εξέταση Θ. Π., Σεπτέμβριος 2008)**

    Επιστήμονες που μελετούν ένα σπάνιο ερπετό έχουν προσδιορίσει ότι η επώαση του κάθε αβγού του ερπετού διαρκεί τουλάχιστον 41 ημέρες και δεν υπερβαίνει τις 104 ημέρες. Κατά τα άλλα, τίποτα δεν είναι γνωστό για την κατανομή της διάρκειας επώασης των αβγών.

    (α) Έστω διακριτή τ.μ. που αναπαριστά τη διάρκεια επώασης ενός αβγού του ερπετού. Δώστε ένα άνω φράγμα και ένα κάτω φράγμα για την εντροπία της , καθώς και τις κατανομές που αντιστοιχούν στο άνω και στο κάτω φράγμα. Υποθέτουμε ότι η επώαση μετράται σε ακέραιο αριθμό ημερών ($X \in \mathbb{N} \cap [41,\ 104]$, δηλαδή δεν μπορεί, για παράδειγμα, να διαρκέσει 65.3 ημέρες).

    (β) Μετά από νέες παρατηρήσεις, επιβεβαιώθηκε ότι $41 \leq X \leq 104$, και προέκυψε, επίσης, ότι η πιθανότητα η επώαση του αβγού να διαρκεί περισσότερο από 56 ημέρες δεν μπορεί να υπερβεί την πιθανότητα η επώαση να διαρκεί 56 ημέρες ή λιγότερο. Επαναλάβετε το Ερώτημα (α) και συγκρίνετε τα νέα φράγματα που υπολογίσατε.

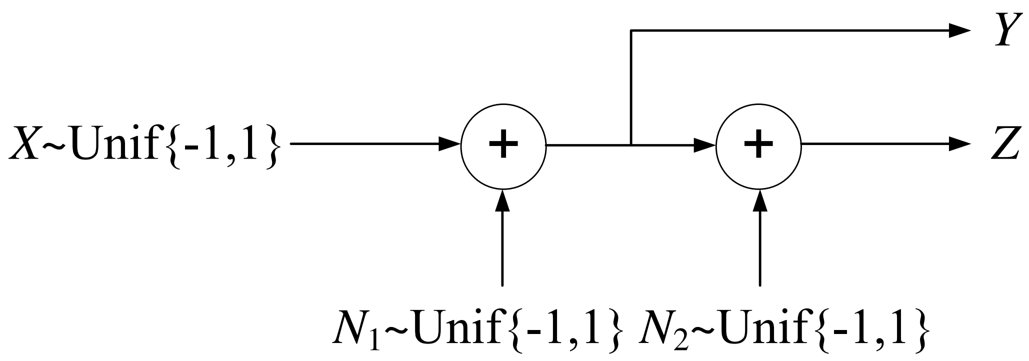
13. **Εντροπία και Αμοιβαία Πληροφορία (Πρόοδος Θ. Π., Νοέμβριος 2008)**

    Στο Σχήμα 1, οι τ.μ. X, N_1 και N_2 είναι διακριτές, δυαδικές και ομοιόμορφα κατανεμημένες. Επίσης, είναι όλες *ανεξάρτητες* μεταξύ τους.

    Στην περίπτωση που σας χρειαστεί, δίνεται ότι $\log_2 3 \approx 1.585$.

    (α) Με τι ισούται η εντροπία της X;

    (β) Βρείτε την εντροπία της Y.



Σχήμα 1: Σύστημα με δύο πηγές θορύβου

(γ) Για *οποιαδήποτε* κατανομή της N_2 (όχι, κατ' ανάγκη ομοιόμορφη όπως στο σχήμα), υποθέτοντας, πάντοτε, ότι η N_2 είναι τυχαία και ανεξάρτητη των X και N_1, δώστε το καλύτερο κάτω φράγμα που μπορείτε για την $H(Z)$. Θεωρούμε ότι η κατανομή της N_2 (και, επομένως, και η εντροπία της) είναι γνωστή.

Προσοχή: Σε αυτό το ερώτημα (αντίθετα με όλα τα άλλα) δε θεωρούμε ότι η N_2 είναι, κατ' ανάγκη, δυαδική.

(δ) Υπολογίστε, τώρα, την $H(Z)$ για $N_2 \sim$ Unif$\{-1, +1\}$ και επαληθεύστε ότι υπερβαίνει το κάτω φράγμα που βρήκατε στο προηγούμενο ερώτημα.

(ε) Υπολογίστε τις $I(Y;X)$, $I(Z;X)$ και $I(Y,Z;X)$. Συγκρίνετέ τις μεταξύ τους και σχολιάστε.

(στ) Σας δίνονται οι εξής επιλογές: Παρατήρηση της τ.μ. Y μόνο, παρατήρηση της τ.μ. Z μόνο ή παρατήρηση του ζεύγους τ.μ. (Y,Z). Ποια παρατήρηση θα επιλέξετε ώστε να πάρετε όση περισσότερη πληροφορία για τη X μπορείτε (κατά μέσο όρο);

14. **Σύστημα με δύο εξόδους (Τελικό διαγώνισμα Θ. Π., Φεβρουάριος 2009)**

Θεωρούμε το σύστημα του Σχήματος 2. Η πηγή *δεν* έχει μνήμη και παίρνει τιμές $+1$ ή -1 με την ίδια πιθανότητα $1/2$. Οι *πολλαπλασιαστικοί* θόρυβοι Z_1 και Z_2 παίρνουν, επίσης, τιμές στο σύνολο $\{-1, 1\}$, αλλά ακολουθούν κατανομή Bernoulli(p_i). Δηλαδή,

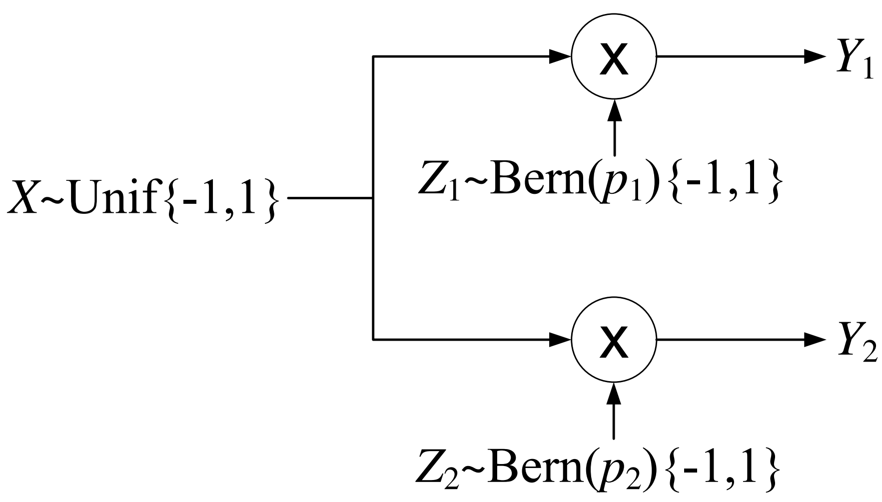
$$Z_i = \begin{cases} +1 & \text{με πιθανότητα} \quad p_i \\ -1 & \text{με πιθανότητα} \quad 1-p_i \end{cases}.$$

Οι τ.μ. Z_1 και Z_2 είναι ανεξάρτητες της X. Ωστόσο, στη γενική περίπτωση, **δεν** είναι ανεξάρτητες μεταξύ τους.

(α) Βρείτε την $H(X)$ και τις $H(Z_i)$ (συναρτήσει των p_i).

(β) Για *δεδομένα* p_1 και p_2, βρείτε ένα άνω φράγμα για την $H(Z_1, Z_2)$, καθώς και την από κοινού συνάρτηση μάζας πιθανότητας $p(z_1, z_2)$ η οποία επιτυγχάνει το άνω φράγμα.

(γ) Βρείτε τις $I(X;Y_1)$ και $I(X;Y_2)$. Για δεδομένα p_1 και p_2 υπάρχει τρόπος να αυξήσετε τις $I(X;Y_1)$ και $I(X;Y_2)$ αλλάζοντας την κατανομή της X; Επιτρέπεται να αλλάξετε μόνο τις πιθανότητες με τις οποίες $X = 1$ ή -1. Δεν μπορείτε να αυξήσετε το πλήθος τιμών, $|\mathcal{X}|$, της πηγής.



Σχήμα 2: Σύστημα με δύο εξόδους.

(δ) Συγκρίνετε την $I(X;Y_1,Y_2)$ με τις $I(X;Y_i)$. Είναι μικρότερη; Μεγαλύτερη; Ίση; Ποια είναι η μεγαλύτερη τιμή (σε bits) που μπορεί να πάρει η $I(X;Y_1,Y_2)$;

Υπόδειξη: Στο ερώτημα αυτό μπορείτε να απαντήσετε χωρίς να βρείτε την τιμή της $I(X;Y_1,Y_2)$.

(ε) Υποθέστε, τώρα, ότι οι Z_1 και Z_2 είναι *ανεξάρτητες* μεταξύ τους. Βρείτε μια έκφραση για την $(X;Y_1,Y_2)$ συναρτήσει των p_1 και p_2.

Υπόδειξη: Ενδέχεται να σας φανεί χρήσιμη η αντικατάσταση $\alpha \triangleq p_1+p_2-2p_1p_2$.

15. **Ανισότητες (Επαναληπτική Εξέταση Θ. Π., Σεπτέμβριος 2009)**

(α) Να αποδειχτεί ότι, για $n \geq 2$,

$$H(X_1,X_2,\ldots,X_n) \geq \sum_{i=1}^{n} H(X_i|X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n).$$

Στη συνέχεια, να αποδειχτεί ότι

$$\frac{1}{2}\left[H(X_1,X_2)+H(X_2,X_3)+H(X_1,X_3)\right] \geq H(X_1,X_2,X_3).$$

Υπόδειξη: Χρησιμοποιήστε τις σχέσεις

$$H(X_1,X_2,X_3) = H(X_i,X_j)+H(X_k|X_i,X_j),$$

καθώς και την πρώτη σχέση που αποδείξατε.

(β) Έστω $p_{XY}(x,y)$ και $q_{XY}(x,y)$ δύο κατανομές μάζας πιθανότητας στο σύνολο $\mathcal{X} \times \mathcal{Y}$. Να αποδειχτεί ότι $D(p_{XY}||q_{XY}) \geq D(p_X||q_X)$, όπου $p_X(x)$ και $q_X(x)$ οι περιθώριες σ.μ.π. των $p_{XY}(x,y)$ και $q_{XY}(x,y)$, αντιστοίχως.

16. **Ισότητες και ανισότητες (Επαναληπτική εξέταση Π. Θ. Θ. Π., Σεπτέμβριος 2009)**

Συμπληρώστε τις παρακάτω σχέσεις με $=$, $\leq$ ή $\geq$. Δικαιολογήστε την απάντησή σας. Στην περίπτωση που ισχύει $\leq$ ή $\geq$, προσδιορίστε πότε ισχύει η ισότητα.

Σημείωση: Απαντήσεις που δεν είναι επαρκώς αιτιολογημένες δε βαθμολογούνται.

(α) $I(X;Y)\ ?\ I(g(X);Y)$. Η $g()$ είναι ντετερμινιστική συνάρτηση.

(β) $I(Y;Z|X)\ ?\ I(Y;Z)$ εάν $p(x,y,z) = p(x)p(y)p(z|x,y)$.

(γ) $H(X|Z)\ ?\ H(X|Y) + H(Y|Z)$.

17. **Μεταβολή εντροπίας (Επαναληπτική Εξέταση Π. Θ. Θ. Π., Σεπτέμβριος 2010)**

Θεωρήστε μια κατανομή $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ με $N \geq 2$ ενδεχόμενα, όλα *μη μηδενικής πιθανότητας*, δηλαδή $p_n > 0\,\forall n$. Θα δείξουμε ότι, για οποιαδήποτε κατανομή, υπάρχει πάντοτε τουλάχιστον ένα ενδεχόμενο, έστω K, το οποίο αν αφαιρέσουμε θα ελαττώσουμε την εντροπία. Δηλαδή, υπάρχει πάντοτε τουλάχιστον ένα p_k έτσι ώστε, αν $\mathbf{p}' = \left(\frac{p_1}{1-p_K}, \frac{p_2}{1-p_K}, \ldots, \frac{p_{K-1}}{1-p_K}, \frac{p_{K+1}}{1-p_K}, \ldots, \frac{p_N}{1-p_K}\right)$, $H(\mathbf{p}') < H(\mathbf{p})$.

(α) Δείξτε ότι δεν μπορούμε πάντοτε να ελαττώσουμε την εντροπία αν επιλέξουμε το ενδεχόμενο τυχαία. Δηλαδή, υπάρχουν κατανομές για τις οποίες υπάρχει p τέτοιο ώστε $H(\mathbf{p}') > H(\mathbf{p})$.

(β) Στη συνέχεια θα αποδείξετε ένα λήμμα που θα σας βοηθήσει να αποδείξετε το ζητούμενο. Δείξτε ότι, για $p \in (0, 1/2]$, η συνάρτηση $\frac{H(p)}{p}$, όπου $H(p) \triangleq -p\log_2 p - (1-p)\log_2(1-p)$ είναι φθίνουσα συνάρτηση του p.

(γ) Εξηγήστε γιατί, σε μία οποιαδήποτε κατανομή με ενδεχόμενα, $(p_1, p_2, \ldots, p_N)$, όλα *μη μηδενικά*, υπάρχει πάντοτε τουλάχιστον ένα ενδεχόμενο, L, με $p_L \leq \frac{1}{N}$.

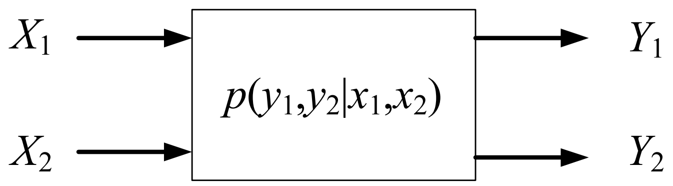
(*δ) Χρησιμοποιώντας τα Ερωτήματα (β) και (γ) (ή κάποιον άλλο τρόπο, αν προτιμάτε) αποδείξτε ότι υπάρχει πάντοτε ένα ενδεχόμενο, K, το οποίο αν αφαιρεθεί η εντροπία της κατανομής μειώνεται.

18. **Ανισότητες (Τελική Εξέταση Θ. Π., Φεβρουάριος 2011)**

Θεωρήστε 4 διακριτές τ.μ. X_1, X_2, Y_1 και Y_2 οι οποίες ικανοποιούν τις εξής συνθήκες

- Οι X_1 και X_2 είναι *ανεξάρτητες*. Δηλαδή, $p_{X_1X_2}(x_1, x_2) = p_{X_1}(x_1) \cdot p_{X_2}(x_2)$.
- Οι Y_1 και Y_2 εξαρτώνται από τις X_1 και X_2 μέσω της $p_{Y_1Y_2|X_1X_2}(y_1, y_2|x_1, x_2)$.

Για παράδειγμα, οι Y_1 και Y_2 μπορεί να είναι έξοδοι του συστήματος του Σχήματος 3, στο οποίο εισάγουμε ανεξάρτητες εισόδους (δηλαδή ο χρήστης 1 δεν μπορεί να συνεννοηθεί με το χρήστη 2). Ωστόσο, στη γενική περίπτωση, η έξοδος Y_j εξαρτάται όχι μόνο από τη X_j, αλλά και από τη X_i, $i \neq j$.



Σχήμα 3: Στοχαστικό σύστημα με εισόδους X_1 και X_2 και εξόδους Y_1 και Y_2.

(α) Δείξτε ότι $I(X_1, X_2; Z) \geq I(X_1; Z) + I(X_2; Z)$, όπου $Z = Y_1$, $Z = Y_2$ ή $Z = (Y_1, Y_2)$.

Δηλαδή, κάποιος που διαθέτει πρόσβαση στα X_1 και X_2 ταυτοχρόνως, μπορεί να αντλήσει περισσότερη πληροφορία για το Z από τη συνολική πληροφορία που μπορούν να αντλήσουν δύο άτομα που διαθέτουν πρόσβαση ο καθένας μόνο στο X_1 και μόνο στο X_2.

Πότε ίσχύει η ισότητα στην παραπάνω σχέση;

(β) Έστω, τώρα, ότι ισχύει $p(y_1, y_2|x_1, x_2) = p(y_1|x_1, x_2) \cdot p(y_2|x_1, x_2)$. Δηλαδή, οι Y_1 και Y_2 είναι ανεξάρτητες *δεδομένων των* X_1 και X_2.

Δείξτε ότι

$$I(X_1, X_2; Y_1, Y_2) \leq I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2).$$

Δηλαδή, κάποιο μέρος της πληροφορίας που περιέχει η Y_1 για τις X_1 και X_2 υπάρχει και στην Y_2 (και αντιστρόφως).

Πότε ισχύει η ισότητα;

19. **Σύγκριση εντροπιών (Τελική Εξέταση Θ. Π., Φεβρουάριος 2012)**

Θεωρούμε *δυαδική* πηγή *χωρίς* μνήμη, η οποία παράγει τα συμβολα 0 και 1 με πιθανότητες p και $1-p$, αντιστοίχως. Έστω ότι συμβολίζουμε την έξοδο της πηγής με X_i.

Για κάθε ζεύγος εξόδων $X_{2k}X_{2k+1}$ της πηγής, ορίζουμε την τ.μ. C_k η οποία ισούται με τον αριθμό των '0' στο ζεύγος $X_{2k}X_{2k+1}$. Δηλαδή, οι πιθανές τιμές της C_k είναι 0, 1 και 2.

(α) Ποια είναι η κατανομή της C_k;

Αν θεωρήσουμε ότι οι C_k είναι έξοδοι μίας πηγής (η οποία παράγει σύμβολα με το μισό ρυθμό σε σχέση με την πηγή που παράγει τις X_i), η πηγή έχει μνήμη;

(β) Βρείτε την εντροπία του *ζεύγους* $(X_{2k}X_{2k+1})$, την εντροπία της C_k και τη διαφορά τους. Σχολιάστε.

Υπόδειξη: Η άσκηση λύνεται γρήγορα αν εφαρμόσετε την αρχή διαχωρισιμότητας της εντροπίας.

(γ) Για ποια τιμή του p μεγιστοποιείται η διαφορά $H(X_{2k}, X_{2k+1}) - H(C_k)$; Υπάρχουν τιμές του p για τις οποίες $H(X_{2k}, X_{2k+1}) = H(C_k)$;

(δ) Επαναλάβετε το Ερώτημα (β) για την περίπτωση τριάδων εξόδων της πηγής. Δηλαδή, ορίζουμε τ.μ. D_k που ισούται με τον αριθμό των '0' στην τριάδα $(X_{3k}, X_{3k+1}, X_{3k+2})$.

(ε) Γενικεύστε για την περίπτωση όπου η τ.μ. E_k ορίζεται ως ο αριθμός των '0' στη n-άδα $(X_{nk}, X_{nk+1}, \ldots, X_{nk+n-1})$. Αρκεί να δώσετε μία (σωστή) γενική σχέση.

Υπόδειξη: $(x+y)^n = x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{2}x^2y^{n-2} + \binom{n}{1}xy^{n-1} + y^n$.

20. **Φράγματα για την εντροπία (Επαναληπτική Εξέταση Θ. Π., Σεπτέμβριος 2012)**

Θεωρούμε διακριτή τ.μ. X με συνάρτηση μάζας πιθανότητας (pmf) **p** $= \{p_1, p_2, \ldots, p_N\}$ και $p_1 \leq p_2 \leq \cdots \leq p_N$.

(α) Δείξτε ότι $H(X) \geq \sum_{n=1}^{N} p_n(1-p_n)$ nats $\geq 1 - p_N$ nats.
Υπόδειξη: Χρησιμοποιήστε την ανισότητα $\ln x \leq x - 1$, $x > 0$.

(β) Δείξτε ότι $H(X) \geq H(p_N)$, όπου $H(p_N) \triangleq p_N \log \frac{1}{p_N} + (1-p_N) \log \frac{1}{1-p_N}$.

Πότε ισχύει η ισότητα;

(γ) Δείξτε ότι $H(X) \geq -\log p_N$.
Πότε ισχύει η ισότητα;

(δ) **Πιο χρονοβόρο**
Δείξτε ότι $H(X) \geq 2(1-p_N)$ bits.
Υπόδειξη: Χρησιμοποιήστε το φράγμα του Ερωτήματος (β) για την περίπτωση όπου $p_N \geq \frac{1}{2}$ και το φράγμα του Ερωτήματος (γ) για την περίπτωση όπου $p_N \leq \frac{1}{2}$.

21. **Μεταβολή εντροπίας (Επαναληπτική Εξέταση Θ. Π., Σεπτέμβριος 2013)**

Θεωρούμε διακριτή κατανομή με μάζες $p_1, p_2, \ldots, p_N$. Έστω ότι η κατανομή δεν είναι ομοιόμορφη, δηλαδή υπάρχουν τουλάχιστον δύο μάζες που δεν είναι ίσες.

Χωρίς βλάβη της γενικότητας, έστω ότι οι μάζες αυτές είναι οι p_1 και p_2. Επίσης, υποθέτουμε ότι $p_1 > p_2$.

Έστω, τώρα, ότι μεταβάλλουμε κατά μία μικρή ποσότητα $\Delta p > 0$ τις p_i ως εξής:

$$\begin{aligned} p_1' &= p_1 - \Delta p \\ p_2' &= p_2 + \Delta p \\ p_i' &= p_i, \ i \neq 1, 2. \end{aligned}$$

Επίσης, θεωρούμε ότι $p_1 - \Delta p \geq p_2 + \Delta p$.

Δηλαδή, κινούμαστε προς την κατεύθυνση της εξίσωσης των p_1 και p_2 κρατώντας τις άλλες μάζες σταθερές.

Δείξτε ότι οποιαδήποτε τέτοια απόπειρα εξίσωσης μαζών κατανομής (δηλαδή οποιαδήποτε κίνηση προς την ομοιόμορφη κατανομή) αυξάνει την εντροπία της κατανομής. Δηλαδή, δείξτε ότι

$$H(p_1', p_2', \ldots, p_N') > H(p_1, p_2, \ldots, p_N).$$

22. **Σύντομες Ερωτήσεις (Τελική Εξέταση Θ. Π., Φεβρουάριος 2014)**

(α) Έστω X διακριτή τ.μ. Συγκρίνετε την $H(X)$ με την $H(X+3)$.

(β) Έστω ότι κωδικοποιούμε *από κοινού* n σύμβολα πηγής $X_1, X_2, \ldots, X_n$ με χρήση κώδικα Huffman.
Αν $c(X_1, X_2, \ldots, X_n)$ είναι η κωδική λέξη που αντιστοιχεί στην ακολουθία $X_1, X_2, \ldots, X_n$, συγκρίνετε την $H(X_1, X_2, \ldots, X_n)$ με την $H\left(c(X_1, X_2, \ldots, X_n)\right)$.

(γ) Στο προηγούμενο ερώτημα, έστω, τώρα, ότι $l(X_1, X_2, \ldots, X_n)$ είναι η συνάρτηση μήκους, δηλαδή το μήκος της κωδικής λέξης $c(X_1, X_2, \ldots, X_n)$. Συγκρίνετε την $H(X_1, X_2, \ldots, X_n)$ με την $H\left(l(X_1, X_2, \ldots, X_n)\right)$.

(δ) Έστω $\{X_n\}$ *στάσιμη* στοχαστική διαδικασία με $X_n \in \mathcal{X}$ και $H(X_n) < \infty$. Διατάξτε τις $H(X_1)$, $H(X_2|X_1)$ και $\frac{1}{2}H(X_1, X_2)$. Δηλαδή αντιστοιχίστε τις 3 ποσότητες στα a, b και c ώστε $a \flat b \sharp c$, όπου τα $\flat$ και $\sharp$ αντιστοιχούν σε $\leq$ ή $<$.

23. **► Ο σκύλος που ψάχνει για το κόκκαλο – Cover & Thomas 4.12**

Ένας σκύλος κινείται επάνω στους ακέραιους αριθμούς. Σε κάθε βήμα συνεχίζει προς την ίδια κατεύθυνση με πιθανότητα 0.9, ενώ αλλάζει κατεύθυνση με πιθανότητα 0.1. Υποθέστε ότι ο σκύλος ξεκινά από τη θέση $X_0 = 0$ και ότι το πρώτο βήμα μπορεί να γίνει προς οποιαδήποτε κατεύθυνση (θετική ή αρνητική) με την ίδια πιθανότητα. Για παράδειγμα, μια πιθανή διαδρομή του σκύλου είναι

$(X_0, X_1, \ldots) = (0, -1, -2, -3, -4, -3, -2, -1, 0, 1, 2, 3, 4, 3, ...)$.

(α) Υπολογίστε την $H(X_0, X_1, \ldots, X_n)$.

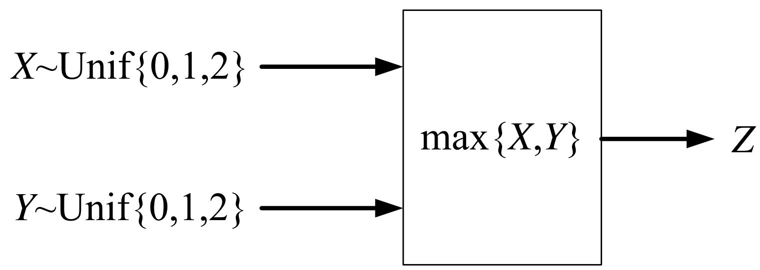
(β) Υπολογίστε το ρυθμό εντροπίας, $H(\mathcal{X})$, της θέσης του σκύλου.

24. **Μέγιστο δύο τ.μ. (Πρόοδος Θ. Π., Νοέμβριος 2009)**

Θεωρούμε το ντετερμινιστικό σύστημα του Σχήματος 4. Οι τ.μ. X και Y ακολουθούν ομοιόμορφη κατανομή στο σύνολο $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$. Οι X και Y είναι *ανεξάρτητες* μεταξύ τους. Το κύκλωμα υπολογίζει την έξοδο με χρήση της σχέσης $Z = \max\{X, Y\}$.

Δίνεται ότι $\log_2 3 \approx 1.585$ και $\log_2 5 \approx 2.3219$.

Εάν προτιμάτε, μπορείτε να αφήσετε στις απαντήσεις σας όρους της μορφής $H(\mathbf{p}) = H(p_1, p_2, \ldots, p_K)$ χωρίς να τους υπολογίσετε, αρκεί να δώσετε τις τιμές των $p_1, p_2, \ldots, p_K$.



Σχήμα 4: Υπολογισμός μεγίστου δύο τ.μ.

(α) Υπολογίστε την $H(X)$ και την $I(X; Y)$.

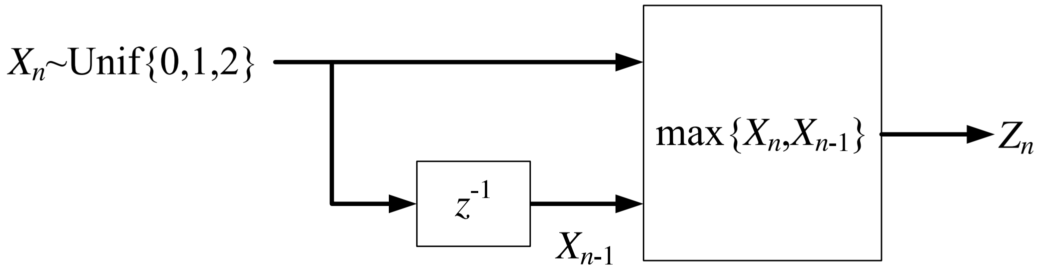
(β) Υπολογίστε την $H(Z)$ και την $I(X; Z)$.

(γ) Βρείτε ένα βέλτιστο *άμεσο δυαδικό* κώδικα για τη Z και συγκρίνετε το μέσο μήκος του με την $H(Z)$.

(δ) Δείξτε ότι $I(X; Y|Z) > 0$ (*όχι* $\geq$). Συγκρίνετε με την $I(X; Y)$ του Ερωτήματος (α).
Υπόδειξη: Η απάντηση προκύπτει πολύ πιο σύντομα αν χρησιμοποιήσετε ιδιότητες εντροπίας ή/και αμοιβαίας πληροφορίας και δείτε τι ισχύει στο συγκεκριμένο πρόβλημα, παρά αν κάνετε πράξεις.

(ε) Εάν τόσο ο συμπιεστής όσο και ο αποσυμπιεστής γνωρίζουν, με κάποιο τρόπο, την τιμή της Y (αλλά όχι της X), προτείνετε ένα βέλτιστο *άμεσο δυαδικό* κώδικα για τη συμπίεση της Z. Πόσο αποδοτικός είναι ο κώδικας;

Θεωρήστε, τώρα, το τροποποιημένο σύστημα του Σχήματος 5. Τώρα χρησιμοποιούμε το κύκλωμα max$\{\cdot\}$ σε διαδοχικές χρονικές στιγμές. Η τ.μ. Y έχει αντικατασταθεί από το δείγμα της X την προηγούμενη χρονική στιγμή, δηλαδή $Z_n = \max\{X_n, X_{n-1}\}$. Θεωρούμε, και πάλι, ότι $X_n \sim$Unif$\{0, 1, 2\}$. Επίσης, οι X_n είναι ανεξάρτητες και ομοίως κατανεμημένες (i.i.d.). Θεωρούμε (παρόλο που δεν έχει ιδιαίτερη σημασία) ότι ο χρόνος αρχίζει τη στιγμή $n = 0$ και ότι $X_{-1} = 0$.



Σχήμα 5: Τροποποιημένο σύστημα.

(στ) Υπολογίστε την $H(Z_n)$ για $n > 0$ και την $I(X_n; X_{n-1})$ για $n \geq 1$.

(*ζ) Υπολογίστε την $H(Z_n|Z_{n-1})$ για $n > 1$. Είναι η $\{Z_n\}$ i.i.d.;

(*η) Συγκρίνετε την $H(Z_n|Z_{n-1}, Z_{n-2})$ (για $n > 2$) με την $H(Z_n|Z_{n-1})$. Δε χρειάζεται να βρείτε ακριβή τιμή, απλώς να προσδιορίσετε εάν
$H(Z_n|Z_{n-1}, Z_{n-2}) = H(Z_n|Z_{n-1})$ ή όχι. Είναι η $\{Z_n\}$ αλυσίδα Markov 1ης τάξης;

25. **Αποταμίευση (Τελική Εξέταση Π. Θ. Θ. Π., Ιούνιος 2009)**

Ένας καταθέτης ανοίγει λογαριασμό με αρχικό κεφάλαιο $X_0 = 1000$ και *μηνιαίο* επιτόκιο 1%. Θεωρούμε ότι το επιτόκιο αυτό είναι εγγυημένο για όσο παραμένει ανοικτός ο λογαριασμός, δηλαδή δε μεταβάλλεται. Επίσης, θεωρούμε ότι ο τόκος υπολογίζεται στο τέλος κάθε μήνα. Στο τέλος κάθε μήνα ο καταθέτης έχει την επιλογή να εισπράξει τον τόκο ή να τον αφήσει στο λογαριασμό, οπότε αυτός προστίθεται στο υπάρχον κεφάλαιο. Θεωρούμε, τέλος, ότι δεν επιτρέπεται στον καταθέτη να εισπράξει ποσό διαφορετικό από τον τόκο στο τέλος κάθε μήνα (ούτε μεγαλύτερο ούτε μικρότερο). Δηλαδή ο καταθέτης πρέπει να εισπράξει είτε τον τόκο του μήνα ή τίποτα.

(α) Εάν σε σύνολο N μηνών ο καταθέτης έχει εισπράξει τον τόκο K φορές, δώστε μια έκφραση για το κεφάλαιο, X_N, στο τέλος του $N-$οστού μήνα. Θεωρούμε ότι η X_N ισούται με το κεφάλαιο που απομένει *μετά* από την είσπραξη του τόκου, εφόσον αυτή γίνει. Εάν ο καταθέτης δεν εισπράξει ποτέ τους τόκους, μετά από πόσους μήνες θα έχει διπλασιάσει το αρχικό κεφάλαιο; Δίνεται ότι $1/\log_2(1.01) \approx 69.66$.

(β) Θεωρούμε, τώρα, ότι ο καταθέτης ενδέχεται να έχει ανάγκη τους τόκους, με αποτέλεσμα να τους εισπράττει στο τέλος κάθε μήνα με πιθανότητα $1/4$. Η απόφαση αν θα εισπράξει τους τόκους το μήνα i είναι ανεξάρτητη από την απόφασή του το μήνα $j \neq i$.

Με τι ισούται η από κοινού εντροπία $H(X_0, X_1, \ldots, X_N)$;

Με τι ισούται ο ρυθμός εντροπίας, $H(\mathcal{X})$;

Με τι ισούται η από κοινού εντροπία $H(X_0, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N)$, για κάποιο $0 < j < N$;

Δίνεται $\log_2 3 \approx 1.585$.

(γ) Δύο φοιτήτριες προσπαθούν να εκτιμήσουν πόσοι μήνες θα χρειαστούν ώστε ο καταθέτης να καταφέρει να *οκταπλασιάσει* το αρχικό του κεφάλαιο. Η εκτίμηση της πρώτης είναι ότι αυτό θα έχει συμβεί σχεδόν σίγουρα σε 209 μήνες. Η δεύτερη ισχυρίζεται ότι η εκτίμηση αυτή είναι παρακινδυνευμένη και υποθέτει ότι θα πρέπει να περιμένουμε τουλάχιστον 279 μήνες. Ποια από τις δύο εκτιμήσεις είναι ορθότερη; Δικαιολογήστε επαρκώς την απάντησή σας. Όπως και στο προηγούμενο ερώτημα, θεωρούμε ότι, στο τέλος κάθε μήνα, ο καταθέτης εισπράττει τους τόκους με πιθανότητα $1/4$.

(*δ) Δώστε μια έκφραση για την τιμή της $H(X_N)$ για $N \to \infty$.

Υπόδειξη: Χρησιμοποιήστε το κεντρικό οριακό θεώρημα. Η έκφραση που θα προκύψει είναι συνάρτηση του N.

26. **Τυπικές ακολουθίες (Επαναληπτική Εξέταση Π. Θ. Θ. Π., Σεπτέμβριος 2010)**

Θεωρούμε πηγή *χωρίς μνήμη* που παράγει ακολουθία δυαδικών τ.μ. $X_i \sim$Bern(p). Δηλαδή, η $X_1, X_2, \ldots$ είναι ακολουθία ανεξάρτητων και ομοίως κατανεμημένων (i.i.d.) τ.μ.

(α) Εάν $\left|A_\epsilon^{(n)}\right|$ είναι το πλήθος των ασθενώς $\epsilon-$τυπικών ακολουθιών μήκους n και $|X_1^n|$ είναι όλες οι δυαδικές ακολουθίες μήκους n, τι μπορούμε να πούμε για το λόγο $\left|A_\epsilon^{(n)}\right| / |X_1^n|$ καθώς $n \to \infty$ και $\epsilon \to 0$;

Υπόδειξη: Πρέπει να δείτε αν ο λόγος συγκλίνει στην ίδια τιμή για όλες τις τιμές της παραμέτρου p.

(β) Θεωρούμε, τώρα, ότι οι τιμές της X_i είναι 0 (με πιθανότητα p) ή 1 (με πιθανότητα $1-p$). Ορίζουμε το βάρος Hamming (Hamming weight), $W(X_1^n)$, της ακολουθίας X_1^n ως τον αριθμό των '1' της ακολουθίας.

Δείξτε ότι $\frac{1}{n}W(X_1^n) \to 1-p$ για $n \to \infty$.

Υπόδειξη: Χρησιμοποιήστε το νόμο των μεγάλων αριθμών.

(γ) Έστω, τώρα, ότι το μήκος, n, της ακολουθίας είναι πεπερασμένο και ότι $\epsilon > 0$. Δείξτε ότι μπορούμε να αποφασίσουμε αν η ακολουθία X_1^n μήκους n είναι *ασθενώς* $\epsilon-$τυπική αν γνωρίζουμε την τιμή του βάρους Hamming, $W(X_1^n)$, της ακολουθίας.

Ενδεικτικές Λύσεις

1. ► **Συναρτήσεις τυχαίων μεταβλητών – Cover & Thomas 2.2 (παραλλαγή)**

(*α) Έστω $y = g(x)$, $g(\cdot)$ ντετερμινιστική συνάρτηση και X μια (διακριτή) τυχαία μεταβλητή. Μπορεί να ειπωθεί κάτι για τη σχέση μεταξύ της εντροπίας της $Y = g(X)$ και της εντροπίας της X; Εάν ναι, τι;

Υπόδειξη: Ο πιο απλός τρόπος να λύσετε την άσκηση είναι με χρήση της αρχής διαχωρισιμότητας της εντροπίας.

Απάντηση:

Από τον ορισμό της εντροπίας, $H(Y) = -\sum_{i=1}^{|\mathcal{Y}|} p(y_i) \log p(y_i)$. Εάν η συνάρτηση $g(X)$ είναι 1-προς-1, σε κάθε x_i αντιστοιχεί ένα y_i και ισχύει $p(x_i) = p(y_i)$ και $|\mathcal{X}| = |\mathcal{Y}|$. Επομένως,

$$H(Y) = -\sum_{i=1}^{|\mathcal{Y}|} p(y_i) \log p(y_i) = -\sum_{i=1}^{|\mathcal{X}|} p(x_i) \log p(x_i) = H(X).$$

Εάν η συνάρτηση $g(\cdot)$ δεν είναι 1-προς-1, υπάρχει τουλάχιστον ένα y_i στο οποίο αντιστοιχούν περισσότερα από ένα $x_{j,i}$ και $p(y_i) = \sum_j p(x_{j,i})$. Άρα, $p(y_i) \geq p(x_{j,i})$ για κάθε j, και, δεδομένου ότι η συνάρτηση λογαρίθμου είναι αύξουσα, $\log p(y_i) \geq \log p(x_{j,i})$ για κάθε j. Επομένως,

$$\begin{aligned}\sum_{x_j:y_i=g(x_j)} p(x_j) \log p(x_j) &\leq \sum_{x_j:y_i=g(x_j)} p(x_j) \log p(y_i) \\ &= \log p(y_i) \sum_{x_j:y_i=g(x_j)} p(x_j) = p(y_i) \log p(y_i).\end{aligned}$$

Εναλλακτικά, δεδομένου ότι οι $p(x_{j,i})/p(y_i)$ για δεδομένο j αποτελούν κατανομή ($\frac{p(x_{j,i})}{p(y_i)} \geq 0$ και $\sum \frac{p(x_{j,i})}{p(y_i)} = 1$), από την ανισότητα Jensen,

$$\begin{aligned}\sum_{x_j:y_i=g(x_j)} \frac{p(x_j)}{p(y_i)} \log p(x_j) &\leq \log \left(\sum_{x_j:y_i=g(x_j)} \frac{p(x_j)}{p(y_i)} p(x_j) \right) \\ &\overset{(i)}{\leq} \log \left(\sum_{x_j:y_i=g(x_j)} p(x_j) \right) = \log p(y_i) \\ \Rightarrow \sum_{x_j:y_i=g(x_j)} p(x_j) \log p(x_j) &\leq p(y_i) \log p(y_i).\end{aligned}$$

(i) $\frac{p(x_{j,i})}{p(y_i)} \leq 1.$

Συνεπώς, μπορούμε να γράψουμε

$$\begin{aligned}H(X) &= -\sum_x p(x) \log p(x) \\ &= -\sum_y \sum_{x:y=g(x)} p(x) \log p(x) \\ &\geq -\sum_y p(y) \log p(y) = H(Y).\end{aligned}$$

Άρα, στη γενική περίπτωση, $H(Y) \leq H(X)$. Το αποτέλεσμα είναι διαισθητικά λογικό. Εάν μια συνάρτηση δεν είναι 1-προς-1, μέρος της πληροφορίας χάνεται κατά την απεικόνιση από τη X στην Y.

Ακόμα πιο απλά, μπορούμε να γράψουμε την $H(X, g(X))$ με δύο τρόπους:

$$H(X, g(X)) = H(X) + H(g(X)|X) = H(g(X)) + H(X|g(X)).$$

Αλλά $H(g(X)|X) = 0$, επειδή η $g(x)$ είναι ντετερμινιστική συνάρτηση και $H(X|g(X)) \geq 0$.

Συνεπώς, $H(X) \geq H(Y) = H(g(X))$.

Όταν η $g(X)$ είναι 1-προς-1 ισχύει, επίσης, $H(X|g(X)) = 0$, οπότε $H(X) = H(Y)$.

(β) Βρείτε τη σχέση μεταξύ $H(X)$ και $H(Y)$ για τις συναρτήσεις $y = x^3$ και $y = \lfloor x/2 \rfloor$, x ακέραιος.

Απάντηση:

Η $g(x) = x^3$ είναι 1-προς-1. Επομένως, $H(Y) = H(X)$. Η $g(x) = \lfloor x/2 \rfloor$ δεν είναι 1-προς-1, στη γενική περίπτωση. Ωστόσο, μπορεί να είναι 1-προς-1 για ορισμένα αλφάβητα. Για παράδειγμα, εάν $X \in \{2, 4, 6\}$, η $\lfloor x/2 \rfloor$ είναι 1-προς-1. Συνεπώς, στη γενική περίπτωση, $H(Y) \leq H(X)$.

2. ► **Εντροπία αθροίσματος τυχαίων μεταβλητών – Cover & Thomas 2.14**

(α) Έστω X και Y τυχαίες μεταβλητές και $Z = X + Y$. Να αποδειχτεί ότι $H(Z|X) = H(Y|X)$.

Απάντηση:

Από τον ορισμό της Δεσμευμένης Εντροπίας,

$$\begin{aligned} H(Z|X) &= \sum_x p(x) H(Z|X = x) \\ &= \sum_x p(x) \sum_z p(Z = z|X = x) \log p(Z = z|X = x). \end{aligned}$$

Εάν γνωρίζουμε ότι $X = x$, $Z = Y + x$. Επομένως, $p_{Z|X}(Z = z|X = x) = p_{Y|X}(Y + x = z|X = x) = p_{Y|X}(Y = z - x|X = x)$. Άρα,

$$\begin{aligned} H(Z|X) &= \sum_x p(x) \sum_z p(Z = z|X = x) \log p(Z = z|X = x) \\ &\overset{(*)}{=} \sum_x p(x) \sum_{y \triangleq z-x} p(Y = z - x|X = x) \log p(Y = z - x|X = x) \\ &= \sum_x p(x) H(Y|X = x) = H(Y|X), \end{aligned}$$

όπου στο $(*)$ χρησιμοποιήσαμε την παρατήρηση ότι, εάν θεωρήσουμε όλες τις πιθανές τιμές της Z για δεδομένη $X = x$, τότε θεωρούμε, ισοδύναμα, όλες τις πιθανές τιμές της Y για $X = x$.

Έχοντας δείξει με βάση τον ορισμό ότι $H(Z|X) = H(Y|X)$, στο μέλλον μπορούμε να γράφουμε σύντομα $H(Z|X) = H(Y + X|X) = H(Y|X)$.

Παρατηρήστε, τέλος, ότι το αποτέλεσμα ισχύει ανεξαρτήτως της ανεξαρτησίας (ή μη) των X και Y.

(*β) Εάν οι X και Y είναι *ανεξάρτητες*, να αποδειχτεί ότι $H(Y) \leq H(Z)$ και $H(X) \leq H(Z)$. Επομένως, όταν σε μια τ.μ. προστίθεται μια ανεξάρτητή της τ.μ., η αβεβαιότητα αυξάνει.

Απάντηση:

Εάν οι X και Y είναι ανεξάρτητες, $H(Y|X) = H(Y)$. Γνωρίζουμε ότι, για οποιεσδήποτε X και Y, $I(X;Y) \geq 0$. Επομένως, $I(Z;X) = H(Z) - H(Z|X) \geq 0 \Rightarrow H(Z) \geq H(Z|X) \stackrel{(a)}{=} H(Y|X) = H(Y) \Rightarrow H(Z) \geq H(Y)$. Με τον ίδιο τρόπο, αποδεικνύεται ότι $H(Z) \geq H(X)$.

Το αποτέλεσμα είναι διαισθητικά λογικό. Δεδομένου ότι οι X και Y είναι ανεξάρτητες, η Z περιέχει περισσότερη πληροφορία από την καθεμία από τις X και Y.

Προσοχή: Δεν ισχύει, κατ' ανάγκη, $H(Z) = H(X) + H(Y)$ (ισχύει, όμως, $H(X,Y) = H(X) + H(Y)$, δεδομένου ότι οι X και Y είναι ανεξάρτητες). Ένας τρόπος να το αποδείξουμε είναι με χρήση του αποτελέσματος της 1ης άσκησης, θεωρώντας τη ντετερμινιστική συνάρτηση $g(X,Y) = X + Y$. Η $g(X,Y)$ είναι, στη γενική περίπτωση, μη αντιστρέψιμη. Επομένως, όπως δείξαμε στην Άσκηση 1, $H(Z) = H(g(X,Y)) \leq H(X,Y) = H(X)+H(Y)$. Διαισθητικά, εάν μας "αποκαλυφθεί" η Z δε μας αποκαλύπτονται, στη γενική περίπτωση, οι ακριβείς τιμές των X και Y. Υπάρχουν, όμως, περιπτώσεις, όπου η $X + Y$ είναι αντιστρέψιμη και $H(Z) = H(X) + H(Y)$ (βρείτε ένα παράδειγμα ως άσκηση).

(*γ) Δώστε ένα παράδειγμα (μη ανεξάρτητων) τ.μ. X και Y για τις οποίες $H(Y) \geq H(Z)$ και $H(X) \geq H(Z)$.

Απάντηση:

Θέλουμε το άθροισμα των X και Y να είναι λιγότερο τυχαίο από τις X και Y. Μία ακραία περίπτωση είναι η $X + Y$ να είναι ντετερμινιστική ποσότητα.

Έστω ότι

$$X = -Y = \begin{cases} +1 & \text{με πιθανότητα} \quad 1/2 \\ -1 & \text{με πιθανότητα} \quad 1/2 \end{cases}$$

$H(X) = H(Y) = H(1/2) = 1$. Ωστόσο, $Z = X + Y = -Y + Y = 0$ και, επομένως, $H(Z) = 0$.

3. **Μια εναλλακτική απόδειξη της ανισότητας $D(p||q) \geq 0$ – Cover & Thomas 2.26**

(α) Δείξτε ότι, για $0 < x < \infty$, $\ln x \leq x - 1$.

Απάντηση:

Η $f(x) = \ln x$ είναι κοίλη $\cap$ για $x > 0$. Συνεπώς, η εφαπτομένη στο $x = 1$ βρίσκεται πάντοτε επάνω από την $f(x)$. Η εφαπτομένη δίνεται από την εξίσωση $y - f(1) = f'(1)(x - 1) \Rightarrow y = \frac{1}{x}|_{x=1}(x - 1) = x - 1$. Επομένως, $\ln x \leq x - 1$.

Εναλλακτικά, το αποτέλεσμα προκύπτει με μεγιστοποίηση της $g(x) = \ln x - x + 1$ (ή με ελαχιστοποίηση της $-g(x)$) και από το γεγονός ότι η $g(x)$ είναι κοίλη $\cap$ στο $(0, +\infty)$ και, επομένως, το τοπικό μέγιστο στο $x = 1$ είναι και ολικό.

(β) Δικαιολογήστε τα παρακάτω βήματα

$$\begin{aligned}-D(p||q) &= \sum_x p(x) \ln \frac{q(x)}{p(x)} \\ &\stackrel{(i)}{\leq} \sum_x p(x) \left(\frac{q(x)}{p(x)} - 1 \right) \\ &\stackrel{(ii)}{\leq} 0.\end{aligned}$$

Απάντηση:

(i) Ερώτημα (α) (ii) Η $q(x)$ ενδέχεται να έχει μη μηδενική μάζα για x όπου $p(x) = 0$, οπότε $\sum_x q(x) \leq 1$.

(γ) Πότε ισχύει η ισότητα;

Απάντηση:

Για να ισχύει η (i) με ισότητα πρέπει $q(x)/p(x) = 1$ για όλα τα x όπου $p(x) > 0$. Προφανώς, στην περίπτωση αυτή ισχύει και η (ii). Συνεπώς, η ισότητα ισχύει εάν και μόνο εάν $\mathbf{p} \equiv \mathbf{q}$.

4. **Ανισότητα Fano – Cover & Thomas 2.32**

Έστω η εξής από κοινού κατανομή μάζας πιθανότητας για τις τ.μ. (X, Y)

		Y		
		a	b	c
X	1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
	2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
	3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Έστω, επίσης, $\hat{X}(Y)$ εκτιμητής για τη X με βάση την Y και $P_e = \Pr\{\hat{X}(Y) \neq X\}$.

(α) Βρείτε το βέλτιστο εκτιμητή $\hat{X}(Y)$, καθώς και την P_e που αντιστοιχεί στο βέλτιστο εκτιμητή.

Απάντηση:

Από τον πίνακα, παρατηρούμε ότι ο βέλτιστος κανόνας απόφασης είναι ο

$$\hat{X}(y) = \begin{cases} 1 & \text{για} \quad y = a \\ 2 & \text{για} \quad y = b \\ 3 & \text{για} \quad y = c \end{cases}$$

Πιο αυστηρά, ο βέλτιστος κανόνας απόφασης (Maximum a Posteriori – MAP – detector) προκύπτει από τη μετιστοποίηση της εκ των υστέρων πιθανότητας

$$\max_x p(x|y).$$

Το βέλτιστο x για κάθε τιμή y της Y προκύπτει απευθείας από τον πίνακα.

Η ελάχιστη πιθανότητα σφάλματος για τον εκτιμητή ισούται με

$$P_e = \text{Pr}\left\{\hat{X}(Y) \neq X\right\} = \sum_x p(x)\,\text{Pr}\left\{\hat{X}(Y) \neq X | X = x\right\}$$
$$= \frac{1}{3}\left\{\text{Pr}\left\{Y \neq a | X = 1\right\} + \text{Pr}\left\{Y \neq b | X = 2\right\} + \text{Pr}\left\{Y \neq c | X = 3\right\}\right\} = \frac{1}{2}.$$

(β) Βρείτε το φράγμα για την P_e που προκύπτει από την ανισότητα Fano και συγκρίνετε.

Απάντηση:

Από την ανισότητα Fano,

$$P_e \geq \frac{H(X|Y) - 1}{\log |\mathcal{X}|}.$$

Η $H(X|Y)$ μπορεί να υπολογιστεί ως εξής

$$H(X|Y) = \sum_y p(y) H(X|Y = y) = \frac{1}{3} \times 3H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) = 1.5 \text{ bits}.$$

Συνεπώς, $P_e \geq \frac{0.5}{\log_2 3} \approx 0.316$.

5. **Σχετική Εντροπία – Cover & Thomas 2.37**

Έστω 3 τ.μ. X, Y, Z με από κοινού συνάρτηση μάζας πιθανότητας $p(x, y, z)$. Η σχετική εντροπία μεταξύ της από κοινού κατανομής και του γινομένου των περιθώριων κατανομών ορίζεται ως

$$D(p(x, y, z)||p(x)p(y)p(z)) = \mathbb{E}\left[\log \frac{p(x, y, z)}{p(x)p(y)p(z)}\right].$$

(α) Εκφράστε την $D(p(x, y, z)||p(x)p(y)p(z))$ συναρτήσει εντροπιών.

Απάντηση:

$$\begin{aligned}
&D(p(x, y, z)||p(x)p(y)p(z)) \\
&= \mathbb{E}_{p_{XYZ}}\left[\log \frac{p(x, y, z)}{p(x)p(y)p(z)}\right] \\
&= \mathbb{E}_{p_{XYZ}}\left[\log \frac{1}{p(x)p(y)p(z)}\right] - \mathbb{E}_{p_{XYZ}}\left[\log \frac{1}{p(x, y, z)}\right] \\
&= \mathbb{E}_{p_{XYZ}}\left[\log \frac{1}{p(x)}\right] + \mathbb{E}_{p_{XYZ}}\left[\log \frac{1}{p(y)}\right] + \mathbb{E}_{p_{XYZ}}\left[\log \frac{1}{p(z)}\right] - \mathbb{E}_{p_{XYZ}}\left[\log \frac{1}{p(x, y, z)}\right] \\
&= H(X) + H(Y) + H(Z) - H(X, Y, Z).
\end{aligned}$$

Παρατηρήστε ότι, δεδομένου ότι $H(X_1, \ldots, X_N) \leq \sum_n H(X_n)$, $D(p(x, y, z)||p(x)p(y)p(z)) \geq 0$, όπως έχουμε αποδείξει για οποιαδήποτε $D(||)$.

(β) Πότε η $D(p(x, y, z)||p(x)p(y)p(z))$ ισούται με 0;

Απάντηση:

Η $D(p(x, y, z)||p(x)p(y)p(z))$ ισούται με 0 όταν $H(X, Y, Z) = H(X) + H(Y) + H(Z)$, δηλαδή εάν και μόνο εάν οι X, Y και Z είναι ανεξάρτητες.

6. **Κέρμα και ζάρι – Cover & Thomas 2.43**

   (α) Θεωρήστε το αποτέλεσμα της ρίψης *αμερόληπτου* κέρματος. Με τι ισούται η αμοιβαία πληροφορία μεταξύ της επάνω και κάτω πλευράς του κέρματος; Συμφωνεί το αποτέλεσμα με αυτό που περιμένατε διαισθητικά;

   Απάντηση:

   Έστω ότι αντιστοιχίζουμε την κορώνα στο $+1$ και τα γράμματα στο -1. Εάν X είναι η τιμή της επάνω πλευράς του κέρματος και Y η τιμή της κάτω πλευράς, $Y = -X$. Επομένως, $I(X;Y) = H(X) - H(X|Y) = H(X) - H(X|-X) = H(X) = H(1/2) = 1$ bit.

   Το αποτέλεσμα συμφωνεί με αυτό που περιμέναμε διαισθητικά: Εάν γνωρίζουμε την επάνω πλευρά του κέρματος, γνωρίζουμε επακριβώς και την κάτω. Επομένως, η "αποκάλυψη" της επάνω πλευράς μας παρέχει όλη την πληροφορία σχετικά με την κάτω πλευρά.

   (β) Αλλάζει η απάντησή σας εάν το κέρμα είναι μεροληπτικό;

   Απάντηση:

   Όχι, εκτός από το ότι $I(X;Y) = H(X) = H(p)$. Το ζητούμενο εδώ δεν είναι η πληροφορία που φέρνει η ρίψη του κέρματος, αλλά η πληροφορία που μας δίνει η μια όψη του κέρματος για την άλλη. Εάν ρίξουμε το κέρμα και δούμε την επάνω του όψη, τότε γνωρίζουμε με ακρίβεια την κάτω ακόμα και όταν το κέρμα είναι μεροληπτικό. Η μόνη περίπτωση να αλλάξει η αμοιβαία πληροφορία είναι με κάποια πιθανότητα το κέρμα να είναι ελαττωματικό και να έχει κορώνα (ή γράμματα) και στις δύο του πλευρές.

   (γ) Θεωρήστε τη ρίψη αμερόληπτου ζαριού με 6 πλευρές. Ποια είναι η αμοιβαία πληροφορία μεταξύ της επάνω και της μπροστινής πλευράς; Υπενθυμίζεται ότι σε ένα ζάρι το άθροισμα των αντίθετων πλευρών ισούται με 7.

   Απάντηση:

   Έστω X η μπροστινή πλευρά και Y η επάνω πλευρά. $I(X;Y) = H(X) - H(X|Y)$. Όταν μας αποκαλύπτεται η μπροστινή πλευρά μαθαίνουμε μόνο την πίσω πλευρά. Επομένως, υπάρχουν 4 εξίσου πιθανά ενδεχόμενα για την επάνω πλευρά. Άρα, $I(X;Y) = \log 6 - \log 4 = \log 3 - 1$ bits.

   Ωστόσο, αν θέλουμε να είμαστε ακόμα πιο σχολαστικοί, πρέπει να παρατηρήσουμε ότι μπορούμε να εκμεταλλευτούμε ακόμα και τον τρόπο με τον οποίο απεικονίζονται οι αριθμοί στο ζάρι. Συγκεκριμένα, το 2, το 3 και το 6 είναι τυπωμένα με τέτοιο τρόπο ώστε να μπορούμε να αποκλείσουμε 2 από τις 4 πλευρές. Επομένως,

   $$H(X|Y) = \frac{1}{2}H(X|Y \in \{2,3,6\}) + \frac{1}{2}H(X|Y \in \{1,4,5\}) = \frac{1}{2}2 + \frac{1}{2}1 = \frac{3}{2},$$

   και $I(X;Y) = \log 6 - \frac{3}{2} = 1 + \log 3 - \frac{3}{2} = \log 3 - \frac{1}{2}$ bits.

7. **Μήκος Ακολουθίας – Cover & Thomas 2.48**

   Θεωρούμε τυχαία διαδικασία Bernoulli$\left(\frac{1}{2}\right)$ $\{X_i\}$. Σταματάμε τη διαδικασία όταν εμφανίζεται το πρώτο "1". Έστω N το μήκος της ακολουθίας όταν σταματάμε. Επομένως, η ακολουθία X^N είναι ένα στοιχείο του συνόλου όλων των δυαδικών ακολουθιών πεπερασμένου μήκους: $\{0,1\}^* = \{0,1,00,01,10,11,000,\ldots\}$.

(α) Βρείτε την $I(N; X^N)$.

(β) Βρείτε την $H(X^N|N)$.

(γ) Βρείτε την $H(X^N)$.

(δ) Βρείτε την $H(N|X^N)$.

(ε) Βρείτε την $H(N)$.

Υπόδειξη: Ενδέχεται να σας φανεί χρήσιμη η Άσκηση 2.1 του βιβλίου των Cover & Thomas.

Απάντηση:

Η κατανομή που ακολουθεί ο αριθμός προσπαθειών είναι η Γεωμετρική. Γενικά, η γεωμετρική κατανομή ορίζεται ως ο αριθμός πειραμάτων Bernoulli έως ότου εμφανιστεί το πρώτο "1" (ή, εναλλακτικά, ως ο αριθμός των αποτυχιών έως ότου εμφανιστεί το "1" – και στις δύο περιπτώσεις η εντροπία είναι η ίδια, όχι, όμως, η μέση τιμή).

$I(N; X^N) = H(N) - H(N|X^N)$. Δεδομένου ότι υπάρχει μόνο μια ακολουθία μήκους N (με $N-1$ "0" και ένα "1" στο τέλος), $H(N|X^N) = 0$. Από τον ορισμό της εντροπίας, εάν $p = \text{Pr}\{X = 1\}$,

$$\begin{aligned}
H(N) &= -\sum_{n=1}^{\infty}(1-p)^{n-1}p \log(1-p)^{n-1}p \\
&= -\sum_{n=1}^{\infty}(1-p)^{n-1}p \log(1-p)^{n-1} - \sum_{n=1}^{\infty}(1-p)^{n-1}p \log p \\
&= -p\sum_{n=0}^{\infty}(1-p)^{n} \log(1-p)^{n} - p\log p\sum_{n=0}^{\infty}(1-p)^{n} \\
&= -p\log(1-p)\sum_{n=0}^{\infty}n(1-p)^{n} - p\log p\sum_{n=0}^{\infty}(1-p)^{n} \\
&\overset{(a)}{=} -p\log(1-p)\frac{1-p}{p^2} - p\log p\frac{1}{p} \\
&= -\frac{(1-p)\log(1-p) - p\log p}{p} = H(p)/p.
\end{aligned}$$

Στο (α) χρησιμοποιήσαμε τους τύπους της Άσκησης 2.1 των Cover & Thomas.

Θέτοντας $p = 2$, $H(N) = 2$ bits. Συνεπώς, $I(N; X^N) = 2$ bits.

$H(X^N|N) = 0$ δεδομένου ότι υπάρχει μόνο μια επιτρεπτή ακολουθία μήκους N. Τέλος, $H(X^N) = H(N)$ γιατί η κατανομή της X^N είναι ίδια με της (σε κάθε N αντιστοιχεί μόνο μια ακολουθία X^N).

Αλλάζουμε, τώρα, τον τρόπο με τον οποίο σταματάμε την παραγωγή της ακολουθίας. Θεωρήστε και πάλι ότι $X_i \sim$Bernoulli$\left(\frac{1}{2}\right)$. Με πιθανότητα $\frac{1}{3}$ σταματάμε όταν $N = 6$, αλλιώς σταματάμε όταν $N = 12$. Επίσης, θεωρούμε ότι η επιλογή της τιμής του N είναι *ανεξάρτητη* της ακολουθίας $X_1X_2 \dots X_{12}$.

(στ) Βρείτε την $I(N; X^N)$.

(ζ) Βρείτε την $H(X^N|N)$.

(η) Βρείτε την $H(X^N)$.

(θ) Βρείτε την $H(N|X^N)$.

(ι) Βρείτε την $H(N)$.

Απάντηση:

$I(N;X^N) = H(N) - H(N|X^N)$. Αν γνωρίζουμε τη X^N γνωρίζουμε και το μήκος της, οπότε $H(N|X^N) = 0$. $H(N) = H(1/3) \approx 0.918$ bits. Επομένως, $I(N;X^N) \approx 0.918$ bits.

$$\begin{aligned} H(X^N|N) &= \text{Pr}\{N=6\}H(X^N|N=6) + \text{Pr}\{N=12\}H(X^N|N=12) \\ &= \frac{1}{3}H(X^6) + \frac{2}{3}H(X^{12}) \\ &\overset{(b)}{=} \frac{1}{3}6H(X) + \frac{2}{3}12H(X) \\ &= 10 \text{ bits.} \end{aligned}$$

(b) Οι X_i είναι i.i.d.

$H(X^N) = I(X^N;N) + H(X^N|N) \approx 10.918$ bits. Εναλλακτικά, από την αρχή διαχωρισιμότητας της εντροπίας,

$$H(X^N) = H(X^N) + H(N|X^N) = H(X^N, N) = H(N) + H(X^N|N) = 10.918 \text{ bits.}$$

Τέλος, η $H(X^N)$ μπορεί να βρεθεί θεωρώντας όλες τις επιτρεπτές ακολουθίες. Υπάρχουν 2^6 ακολουθίες μήκους 6, η κάθε μία με πιθανότητα εμφάνισης $\frac{1}{3}2^{-6}$ και 2^{12} ακολουθίες μήκους 12, η κάθε μία με πιθανότητα εμφάνισης $\frac{2}{3}2^{-12}$.

8. **Εντροπία και Αμοιβαία Πληροφορία (Πρόοδος Θ. Π., Νοέμβριος 2007)**

Έστω οι *ανεξάρτητες* τ.μ. X και Y. Η X ακολουθεί ομοιόμορφη κατανομή με τιμές $\{-1, 1\}$ (δηλαδή $X = 1$ ή $X = -1$ με την ίδια πιθανότητα), ενώ η Y ακολουθεί ομοιόμορφη κατανομή με *γνωστές* τιμές $\{-a, a\}$, $a \in \mathbb{R}$ (δηλαδή $Y = a$ ή $Y = -a$ με την ίδια πιθανότητα). Η τ.μ. Z δίνεται από τη σχέση $Z = X + Y$. Για όλες τις πιθανές τιμές του $a \in \mathbb{R}$ να βρεθούν

(α) Η $H(Y)$.

Απάντηση:

Εάν $a \neq 0$, $H(Y) = \log 2 = 1$ bit (δεδομένου ότι η κατανομή της είναι ομοιόμορφη), αλλιώς, εάν $a = 0$, $H(Y) = 0$ bits.

(β) Η $H(Z)$.

Απάντηση:

• Εάν $a = 0$, $Z = X$. Επομένως, $H(Z) = H(X) = 1$ bit.

• Εάν $a = \pm 1$, η κατανομή της Z είναι $(-2, 0, 2)$ με πιθανότητες $\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$, αντιστοίχως. Επομένως, $H(Z) = 2 \times \frac{1}{4}\log 4 + \frac{1}{2}\log 2 = 1.5$ bits.

• Για όλα τα άλλα a, η Z ακολουθεί ομοιόμορφη κατανομή στο σύνολο $\{1+a, 1-a, -1+a, -1-a\}$. Συνεπώς, $H(Z) = \log 4 = 2$ bits.

(γ) Η $I(X;Z)$.

Απάντηση:

$I(X;Z) = H(Z) - H(Z|X)$.

• Εάν $a = 0$, $Z = X$. Επομένως, $H(Z|X) = H(X|X) = 0$ και $I(X;Z) = H(Z) = 1$ bit.

• Εάν $a = \pm 1$: Εάν γνωρίζουμε τη X η Z μπορεί να πάρει 2 τιμές με την ίδια πιθανότητα. Επομένως, $H(Z|X) = 1$ bit και $I(X;Z) = 1.5 - 1 = 0.5$ bits.

• Για όλα τα άλλα a, εάν γνωρίζουμε τη X η Z μπορεί να πάρει 2 τιμές με την ίδια πιθανότητα. Επομένως, $H(Z|X) = 1$ bit και $I(X;Z) = 2 - 1 = 1$ bit.

Εναλλακτικά, $I(X;Z) = H(X) - H(X|Z)$. Εάν γνωρίζουμε τη Z, μπορούμε να βρούμε και τη X εκτός από την περίπτωση όπου $a = \pm 1$. Συνεπώς, για $a \neq \pm 1$, $H(X|Z) = 0$ και $I(X;Z) = H(X) = 1$ bit.

Εάν $a = \pm 1$ και $Z = 2$ ή -2, τότε και πάλι μπορούμε να βρούμε τη X με βεβαιότητα και, επομένως, $H(X|Z = \pm 2) = 0$. Εάν $Z = 0$ τότε $X = -1$ ή 1 με την ίδια πιθανότητα. Άρα, $H(X|Z = 0) = 1$ bit. Επομένως, $H(X|Z) = \Pr\{Z = \pm 2\} H(X|Z = \pm 2) + \Pr\{Z = 0\} H(X|Z = 0) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = 0.5$ bits, και $I(X;Z) = H(X) - H(X|Z) = 1 - 0.5 = 0.5$ bits.

Παρατηρήσεις:

Ο βασικός σκοπός της άσκησης ήταν να παρατηρήσετε ότι εάν το a ισούται με ± 1 η Z παίρνει 3 αντί για 4 διακριτές τιμές, με αποτέλεσμα η εντροπία της να μην ισούται με το άνω φράγμα (2 bits).

Ένα λάθος που έκαναν αρκετοί ήταν να προσθέσουν την εντροπία της X και της Y για να βρουν την εντροπία της Z. Στο μάθημα δείξαμε ότι, εάν οι X και Y είναι ανεξάρτητες, τότε $H(X,Y) = H(X) + H(Y)$ (η από κοινού εντροπία) και όχι, κατ' ανάγκη, $H(X + Y) = H(X) + H(Y)$ (βλ. Άσκηση 2.14 των Cover & Thomas). Συνεπώς, ενώ $H(X,Y) = H(X) + H(Y)$ ανεξαρτήτως του a, η ισότητα $H(X+Y) = H(X) + H(Y)$ ισχύει μόνο όταν η $g(X,Y) = X + Y$ είναι 1-προς-1 συνάρτηση της (X,Y), δηλαδή όταν σε κάθε τιμή $X + Y$ αντιστοιχεί μοναδικό ζεύγος (X,Y). Στο συγκεκριμένο πρόβλημα η $X + Y$ είναι 1-προς-1 μόνο όταν $a \neq \pm 1$. Εάν $a = \pm 1$, τα ζεύγη $(X,Y) = (+1,-1)$ και $(X,Y) = (-1,+1)$ απεικονίζονται στην ίδια τιμή $X + Y = 0$, με αποτέλεσμα η απεικόνιση να μην είναι 1-προς-1.

9. **Εντροπία, Αμοιβαία Πληροφορία και συμπίεση (Τελικό διαγώνισμα Θ. Π., Φεβρουάριος 2008)**

Έστω οι *ανεξάρτητες* τ.μ. X και Y, η καθεμία με ομοιόμορφη κατανομή στο σύνολο $\{0,1,2,3\}$. Η τ.μ. Z δίνεται από τη σχέση $Z = X \cdot Y$ (γινόμενο). Να βρεθούν

(α) Η $H(X)$.

Απάντηση:

Δεδομένου ότι η X ακολουθεί ομοιόμορφη κατανομή με 4 τιμές, $H(X) = \log 4 = 2$ bits.

(β) Η $H(Z)$. $\log_2 7 \approx 2.8074$.

Απάντηση:

Η Z ισούται με 0 όταν $(X, Y) = (0, *)$ ή όταν $(X, Y) = (*, 0)$ (συνολικά 7 από 16 ισοπίθανα ενδεχόμενα). Με παρόμοιο τρόπο μπορούμε να υπολογίσουμε την κατανομή της Z: $Z = (0, 1, 2, 3, 4, 6, 9)$ με πιθανότητα $\left(\frac{7}{16}, \frac{1}{16}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{8}, \frac{1}{16}\right)$. $H(Z) = \sum p_i \log \frac{1}{p_i} \approx 2.397$ bits.

(γ) Η $I(X; Z)$.

Απάντηση:

$I(X; Z) = H(Z) - H(Z|X)$. Εάν $X = 0$, τότε και $Z = 0$. Επομένως, $H(Z|0) = 0$. Εάν $X = 1, 2$ ή 3 μπορούμε να γράψουμε $p(Z|X \neq 0) = p(X \cdot Y|X \neq 0) = p(Y|X \neq 0) = p(Y)$. Συνεπώς για $X \neq 0$, $H(Z|X) = H(Y) = 2$ bits. Άρα, $H(Z|X) = \frac{1}{4}\left(H(Z|0) + 3H(Z|X \neq 0)\right) = 1.5$ bits και $I(X; Z) \approx 2.397 - 1.5 = 0.897$ bits.

(δ) Κατασκευάστε ένα <u>δυαδικό</u> κώδικα Huffman για την τ.μ. Z. Συγκρίνετε το μέσο μήκος κώδικα με την εντροπία της Z και σχολιάστε.

Απάντηση:

Ακολουθούμε την τυπική διαδικασία για να κατασκευάσουμε ένα κώδικα Huffman. Σημειώνεται ότι υπάρχουν περισσότεροι από ένας κώδικες Huffman (όλοι, όμως, με το ίδιο μέσο μήκος).

0	$\frac{7}{16}$	$\frac{7}{16}$	$\frac{7}{16}$	$\frac{7}{16}$	$\boxed{1}$
2	$\frac{1}{8}$	$\boxed{\frac{1}{4}}$	$\frac{1}{4}$	$\boxed{\frac{9}{16}}$	
3	$\frac{1}{8}$	$\boxed{\frac{3}{16}}$	$\boxed{\frac{5}{16}}$		
6	$\frac{1}{8}$	$\boxed{\frac{1}{8}}$			
1	$\frac{1}{16}$				
4	$\frac{1}{16}$				
9	$\frac{1}{16}$				

Ο κώδικας που προκύπτει είναι ο

$$
\begin{array}{ccc}
0 & \longleftrightarrow & 0 \\
1 & \longleftrightarrow & 1101 \\
2 & \longleftrightarrow & 100 \\
3 & \longleftrightarrow & 101 \\
4 & \longleftrightarrow & 1110 \\
6 & \longleftrightarrow & 1100 \\
9 & \longleftrightarrow & 1111
\end{array}
$$

Το μέσο μήκος του κώδικα ισούται με $\mathbb{E}[l] = \frac{7}{16} \times 1 + 2 \times \frac{1}{8} \times 3 + \left(\frac{1}{8} + 3 \times \frac{1}{16}\right) \times 4 \approx 2.4375$ bits. Παρατηρούμε ότι ο κώδικας Huffman επιτυγχάνει συμπίεση με απόσταση μόλις ~ 0.04 bits μακριά από την εντροπία.

$\boxed{\text{Παρατηρήσεις:}}$

Στην άσκηση αυτή οι επιδόσεις ήταν, γενικά, καλές. Στο Ερώτημα (γ), κάποιοι προτίμησαν να υπολογίσουν την $I(X;Z)$ με χρήση της $H(X)-H(X|Z)$. Η προσέγγιση αυτή, αν και σωστή, ήταν πιο χρονοβόρα. Επίσης, κάποιοι έκαναν λάθος στην εξής λεπτομέρεια: Εάν $Z=0$, η X δεν είναι ομοιόμορφα κατανεμημένη. Από τα 16 ισοπίθανα ενδεχόμενα της μορφής (X,Y), 7 οδηγούν σε $Z=0$. Από αυτά, τα 4 είναι της μορφής $(0,Y)$. Επομένως, είναι πιο πιθανό η έξοδος $Z=0$ να οφείλεται σε $X=0$, παρά σε $X=1$, 2 ή 3. Μαθηματικά, $p(X|Z=0)=\frac{p(X,Z=0)}{p(Z=0)}=\frac{16}{7}p(X,Z=0)=\frac{16}{7}p(X)p(Z=0|X)$. Αντικαθιστώντας, $p(X|Z=0)=4/7$ για $X=0$ και 1/7 για τις άλλες τιμές του X. Επομένως, $H(X|Z=0)=1.6645$ bits και όχι 2 bits.

Ένα άλλο λάθος που έκαναν κάποιοι ήταν να γράψουν το εξής: $H(Z|X)=H(X\cdot Y|X)=H(Y|X)=H(Y)=2$ bits. Το λάθος εδώ είναι ότι, στην περίπτωση που $X=0$, η $X\cdot Y$ δεν είναι αντιστρέψιμη, επομένως, $Y\neq Z/X$.

10. **Σύντομες Ερωτήσεις (Τελικό Διαγώνισμα Θ. Π., Φεβρουάριος 2008)**

   (α) Αποδείξτε ότι $H(X,Y,Z)-H(X,Y)\leq H(X,Z)-H(X)$. Πότε ισχύει η ισότητα;

   $\boxed{\text{Απάντηση:}}$

   Από τον κανόνα αλυσίδας για την εντροπία, $H(X,Y,Z)=H(X,Y)+H(Z|X,Y)\Rightarrow H(X,Y,Z)-H(X,Y)=H(Z|X,Y)$. Ομοίως, $H(X,Z)-H(X)=H(Z|X)$. Δεδομένου ότι (όπως έχουμε δείξει στο μάθημα) η υπό συνθήκη εντροπία δεν υπερβαίνει την εντροπία, $H(Z|X,Y)\leq H(Z|X)\Rightarrow H(X,Y,Z)-H(X,Y)\leq H(X,Z)-H(X)$. Η ισότητα ισχύει όταν $H(Z|X,Y)\leq H(Z|X)$, όταν, δηλαδή, οι Y και Z είναι <u>υπό συνθήκη</u> ανεξάρτητες δεδομένης της X. Για όσους προτιμούν μεγαλύτερη μαθηματική αυστηρότητα, $H(Z|X)-H(Z|X,Y)=I(X;Y|Z)\geq 0$, με $=$ όταν $I(Y;Z|X)=0$.

   (β) Σωστό ή λάθος; $I(X;Y|Z)=0\Rightarrow I(X;Y)=0$. Αποδείξτε ή δώστε αντιπαράδειγμα.

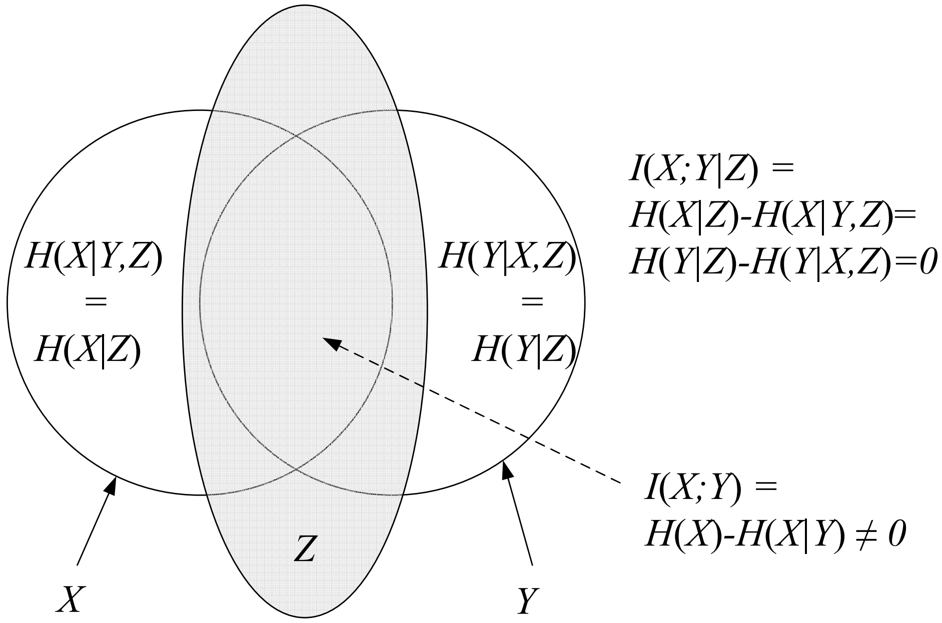
   $\boxed{\text{Απάντηση:}}$

   $\boxed{\text{Λάθος.}}$ Υπό συνθήκη ανεξαρτησία δε συνεπάγεται και ανεξαρτησία. Έστω, για παράδειγμα, $Z=X$ και $Y=X+N$. $I(X;Y|Z)=H(X|Z)-H(X|Y,Z)=H(X|X)-H(X|X,X+N)=0$. Ωστόσο, $I(X;Y)=H(X)-H(X|Y)=H(X)-H(X|X+N)=H(X)-H(N)\neq 0$, στη γενική περίπτωση. Εναλλακτικά, το αναληθές της πρότασης μπορεί να δειχτεί και με χρήση διαγραμμάτων Venn (βλ. Σχήμα 6).

11. **Εντροπία (Τελικό Διαγώνισμα Π. Θ. Θ. Π., Ιούνιος 2008)**

   Θεωρούμε μια διακριτή τυχαία τ.μ. που παίρνει τιμές από ένα πεπερασμένο σύνολο στοιχείων $\mathcal{X}=\{1,\ldots,|\mathcal{X}|\}$. Έστω ότι δίνεται η $p(1)=\Pr\{X=1\}$ η οποία δεν είναι δυνατόν να μεταβληθεί.

   (α) Ποιες είναι οι τιμές $p(i)$, $i=2,\ldots,|\mathcal{X}|$, της κατανομής που μεγιστοποιεί την εντροπία $H(X)$ (για δεδομένη $p(1)$);



Σχήμα 6: Εναλλακτική απάντηση στο Ερώτημα (γ)

Απάντηση:

Από τον ορισμό της εντροπίας,

$$H(X) = -\sum_{i=1}^{|\mathcal{X}|} p(i) \log p(i) = -p(1) \log p(1) - \sum_{i=2}^{|\mathcal{X}|} p(i) \log p(i).$$

Ο δεύτερος όρος μεγιστοποιείται επιλέγοντας ομοιόμορφη κατανομή $p(i) = (1 - p(1))/(|\mathcal{X}| - 1)$, $i = 2, \ldots, |\mathcal{X}|$.

(β) Με τι ισούται η μέγιστη $H(X)$ για δεδομένη $p(1)$; Δώστε μια ερμηνεία της έκφρασης για την $H(X)$ με χρήση της αρχής διαχωρισιμότητας.

Υπόδειξη: Χωρίστε το $\mathcal{X}$ σε δύο κατάλληλα υποσύνολα.

Απάντηση:

Αντικαθιστώντας,

$$\begin{aligned} H(X) &= -p(1) \log p(1) - \sum_{i=2}^{|\mathcal{X}|} \frac{1 - p(1)}{|\mathcal{X}| - 1} \log \frac{1 - p(1)}{|\mathcal{X}| - 1} \\ &= -p(1) \log p(1) - (1 - p(1)) \log \frac{1 - p(1)}{|\mathcal{X}| - 1} \\ &= -p(1) \log p(1) - (1 - p(1)) \log(1 - p(1)) + (1 - p(1)) \log(|\mathcal{X}| - 1) \\ &= H(p(1)) + (1 - p(1)) \log(|\mathcal{X}| - 1). \end{aligned}$$

Το αποτέλεσμα θα μπορούσε να προκύψει και με χρήση του κανόνα αλυσίδας (διαχωρισιμότητα της εντροπίας). Έστω η τ.μ. :

$$Z = \begin{cases} 0 & \text{εάν} \quad X = 1 \\ 1 & \text{εάν} \quad X \neq 1. \end{cases}$$

Εάν χωρίσουμε το σύνολο $\mathcal{X}$ σε δύο υποσύνολα $\{1\}$ και $\{2, 3, \ldots, |\mathcal{X}|\}$, η Z περιγράφει σε ποιο σύνολο ανήκει η X. Θεωρούμε, δηλαδή, ότι, αντί να αποκαλύπτεται απευθείας η τιμή X, αποκαλύπτεται πρώτα το υποσύνολο Z στο οποίο ανήκει η X και, στη συνέχεια, η τιμή της. Μπορούμε να γράψουμε

$$\begin{aligned} H(X) &\overset{(i)}{=} H(X,Z) \overset{(ii)}{=} H(Z) + H(X|Z) \\ &= H(p(1)) + p(1)H(X|Z=0) + (1-p(1))H(X|Z=1) \\ &\overset{(iii)}{=} H(p(1)) + (1-p(1))\,\mathsf{log}(|\mathcal{X}|-1). \end{aligned}$$

(i) $H(X,Z) = H(X) + H(Z|X) = H(X)$, (ii) κανόνας αλυσίδας για την εντροπία, (iii) η εντροπία τ.μ. X δεδομένου ότι $Z = 1$ μεγιστοποιείται από την ομοιόμορφη κατανομή $p(x|Z=1) = 1/(\mathsf{log}\,|\mathcal{X}|-1)$.

(γ) Εάν ένας παίκτης Α μπορεί να μεταβάλλει την $p(1)$ με σκοπό να μειώνει την εντροπία $H(X)$ όσο περισσότερο μπορεί, ενώ ένας παίκτης Β μπορεί να μεταβάλλει οποιοδήποτε υποσύνολο των υπόλοιπων $p(i)$ (αλλά όχι την $p(1)$) με σκοπό να αυξάνει την εντροπία όσο μπορεί, ποια τιμή πρέπει να επιλέξει ο Α για την $p(1)$ αν παίζει πρώτος; Πώς πρέπει να απαντήσει ο Β; Αλλάζει η απάντησή σας εάν πρώτος παίζει ο Β; Θεωρούμε ότι οι παίκτες παίζουν έτσι ώστε να μην παραβιάζεται η συνθήκη $\sum_{i=1}^{|\mathcal{X}|} p(x_i) = 1$ για το **p**.

Απάντηση:

Όταν ο Α παίζει πρώτος, μπορεί να επιλέξει $p(1) = 1$. Επομένως, ο Β είναι αναγκασμένος να επιλέξει $p(i) = 0$, $i = 2, \ldots, |\mathcal{X}|$ και $H(X) = 0$. Εάν ο Β παίζει πρώτος, πρέπει να επιλέξει $p(i) = 1/|\mathcal{X}|$, $i = 2, \ldots, |\mathcal{X}|$, αναγκάζοντας τον Α να θέσει $p(1) = 1/|\mathcal{X}|$. Με τον τρόπο αυτό επιτυγχάνεται μέγιστη εντροπία $H(X) = \mathsf{log}\,|\mathcal{X}|$. Επομένως, η επιλογή του παίκτη που ξεκινά έχει καθοριστική σημασία για την έκβαση του παιχνιδιού.

12. **Εντροπία (Επαναληπτική εξέταση Θ. Π., Σεπτέμβριος 2008)**

Επιστήμονες που μελετούν ένα σπάνιο ερπετό έχουν προσδιορίσει ότι η επώαση του κάθε αβγού του ερπετού διαρκεί τουλάχιστον 41 ημέρες και δεν υπερβαίνει τις 104 ημέρες. Κατά τα άλλα, τίποτα δεν είναι γνωστό για την κατανομή της διάρκειας επώασης των αβγών.

(α) Έστω διακριτή τ.μ. που αναπαριστά τη διάρκεια επώασης ενός αβγού του ερπετού. Δώστε ένα άνω φράγμα και ένα κάτω φράγμα για την εντροπία της , καθώς και τις κατανομές που αντιστοιχούν στο άνω και στο κάτω φράγμα. Υποθέτουμε ότι η επώαση μετράται σε ακέραιο αριθμό ημερών ($X \in \mathbb{N} \cap [41,\ 104]$, δηλαδή δεν μπορεί, για παράδειγμα, να διαρκέσει 65.3 ημέρες).

Απάντηση:

Η ελάχιστη εντροπία της X ισούται με 0 και αντιστοιχεί στην περίπτωση που ο χρόνος επώασης είναι ίδιος για όλα τα αβγά, δηλαδή $p(i) = 1$ για κάποιο i, $40 < i \leq 104$ και $p(j) = 0$ για όλα τα $j \neq i$. Η $H(X)$ μεγιστοποιείται όταν όλα τα ενδεχόμενα είναι ισοπίθανα, όταν, δηλαδή, $p(i) = \frac{1}{64}$, $40 < i \leq 104$ και $H(X) = \mathsf{log}_2 64 = 6$ bits. Επομένως, $\boxed{0 \leq H(X) \leq 6 \text{ bits.}}$

(β) Μετά από νέες παρατηρήσεις, επιβεβαιώθηκε ότι $41 \leq X \leq 104$, και προέκυψε, επίσης, ότι η πιθανότητα η επώαση του αβγού να διαρκεί περισσότερο από 56 ημέρες δεν μπορεί να υπερβεί την πιθανότητα η επώαση να διαρκεί 56 ημέρες ή λιγότερο. Επαναλάβετε το Ερώτημα (α) και συγκρίνετε τα νέα φράγματα που υπολογίσατε.

Δίνεται ότι $\log_2 3 \approx 1.585$.

Απάντηση:

Όπως και στην προηγούμενη περίπτωση, η ελάχιστη τιμή της $H(X)$ ισούται με 0 και επιτυγχάνεται με $p(i) = 1$ για κάποιο i, τέτοιο ώστε $40 < i \leq 56$. Η κατανομή ικανοποιεί τη συνθήκη Pr$\{X > 56\} \leq$ Pr$\{X \leq 56\}$, δεδομένου ότι Pr$\{X \leq 56\} = 1$.

Προκειμένου να υπολογίσουμε ένα άνω φράγμα για την $H(X)$, υποθέτουμε κατ' αρχάς, ότι Pr$\{X \leq 56\} = p$. Από τον περιορισμό, $p \geq 1/2$. Έστω τ.μ. που ορίζεται ως

$$Y = \begin{cases} 0 & \text{εάν} \quad 40 < X \leq 56 \\ 1 & \text{εάν} \quad 56 < \ \leq 104 \end{cases}.$$

$H(X,Y) = H(X) + H(Y|X) = H(X)$. Επίσης, $H(X,Y) = H(Y) + H(X|Y)$. Συνεπώς, $H(X) = H(Y) + H(X|Y)$ (αρχή διαχωρισιμότητας εντροπίας). Αντικαθιστώντας, $H(X) = H(p) + pH(X|Y = 0) + (1-p)H(X|Y = 1)$, όπου $H(p) \triangleq -p \log p - (1-p)\log(1-p)$.

Η $H(X|Y = 0)$ μεγιστοποιείται για ομοιόμορφη κατανομή $p(i|i \leq 56) = \frac{1}{16}$ ανεξαρτήτως της τιμής της p, και $H(X|Y = 0) = \log_2 16 = 4$ bits. Η $H(X|Y = 1)$ μεγιστοποιείται για ομοιόμορφη κατανομή $p(i|i > 56) = \frac{1}{48}$, και $H(X|Y = 1) = \log_2 48 = \log_2(3 \cdot 16) = 4 + \log 3 \approx 5.585$ bits.

Συνεπώς, $H(X) = -p\log p - (1-p)\log(1-p) + 4p + (4+\log 3)(1-p) = -p\log p - (1-p)\log(1-p) - p\log 3 + \log 48$. Η $H(X)$ είναι κοίλη ($\cap$) συνάρτηση δεδομένου ότι είναι άθροισμα μιας κοίλης συνάρτησης $H(p)$ και μιας συνάρτησης που είναι γραμμική ως προς p ($-p\log 3 + \log 48$). Εναλλακτικά, μπορείτε να το επιβεβαιώσετε υπολογίζοντας τη δεύτερη παράγωγο της $H(X)$ (η δεύτερη παράγωγος του όρου $-p\log 3 + \log 48$ ισούται με 0). Επομένως, μπορούμε να βρούμε την τιμή του $p \in [\frac{1}{2}, 1]$ που μεγιστοποιεί την $H(X)$ παραγωγίζοντας ως προς p.

$$\frac{\partial H(X)}{\partial p} = -\log_2 p - \log_2 e + \log_2(1-p) + \log_2 e - log_2 3 = \log_2 \frac{1-p}{3p}.$$

Η $H(X)$ μηδενίζεται όταν $1-p = 3p \Rightarrow p = \frac{1}{4}$. Για $p > \frac{1}{4}$, $\frac{\partial H(X)}{\partial p} < 0$. Επομένως, η τιμή του $p \in [\frac{1}{2}, 1]$ για την οποία μεγιστοποιείται η $H(X)$ είναι $p = \frac{1}{2}$ και $H(X) = 1 - \frac{1}{2}\log 3 + \log 48 \approx 5.7925$ bits. Συνεπώς, $\boxed{0 \leq H(X) \leq 5.7925 \text{ bits.}}$

Εναλλακτική λύση:

Δεδομένου ότι η είναι ομοιόμορφη στα διαστήματα $[41, 56]$ και $[57, 104]$,

$$p(i) = \begin{cases} \frac{p}{16} & , \quad 40 < i \leq 56 \\ \frac{(1-p)}{48} & , \quad 56 < i \leq 104 \end{cases}$$

Επομένως, $H(X) = -16 \cdot \frac{p}{16} \log_2 \frac{p}{16} - 48 \cdot \frac{1-p}{48} \log_2 \frac{1-p}{48} = H(p) + 4p + (1-p)\log_2 48$. Η έκφραση είναι η ίδια με αυτήν που προκύπτει από την αρχή διαχωρισιμότητας. Συνεπώς, συνεχίζουμε όπως προηγουμένως.

Παρατηρήστε ότι το άνω φράγμα για την εντροπία είναι μικρότερο σε αυτήν την περίπτωση. Ο λόγος είναι ότι τώρα γνωρίζουμε κάτι περισσότερο, ότι, δηλαδή, Pr$\{X > 56\} \leq$ Pr$\{X \leq 56\}$. Επομένως, η αβεβαιότητά μας για την κατανομή της X (και άρα την εντοπία της X) ελαττώνεται.

Επίσης, παρατηρήστε ότι, για $p = 1/4$, $H(X) = 6$ bits, που αντιστοιχεί στην περίπτωση του Ερωτήματος (α).

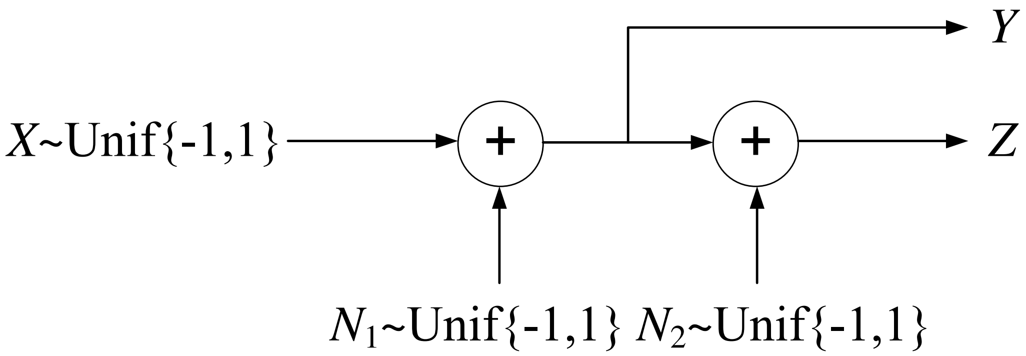
Παρατηρήσεις:

Όσοι απάντησαν στο Ερώτημα (β) υπέθεσαν ότι Pr$\{X > 56\} =$ Pr$\{X \leq 56\} = \frac{1}{2}$. Η τιμή αυτή μεγιστοποιεί, πράγματι, την εντροπία, αλλά αυτό πρέπει να αποδειχτεί. Επίσης, παρατηρήστε ότι, εάν Pr$\{X > 56\} =$ Pr$\{X \leq 56\} = \frac{1}{2}$, το κάτω φράγμα για την $H(X)$ δεν είναι 0, αλλά 1 bit.

13. **Εντροπία και Αμοιβαία Πληροφορία (Πρόοδος Θ. Π., Νοέμβριος 2008)**

Στο Σχήμα 7, οι τ.μ. X, N_1 και N_2 είναι διακριτές, δυαδικές και ομοιόμορφα κατανεμημένες. Επίσης, είναι όλες *ανεξάρτητες* μεταξύ τους.

Στην περίπτωση που σας χρειαστεί, δίνεται ότι $\log_2 3 \approx 1.585$.



Σχήμα 7: Σύστημα με δύο πηγές θορύβου

(α) Με τι ισούται η εντροπία της X;

Απάντηση:

Η ακολουθεί ομοιόμορφη κατανομή με 2 τιμές. Επομένως, η εντροπία της ισούται με 1 bit. Το ίδιο αποτέλεσμα προκύπτει και με απευθείας υπολογισμό της $H(X)$.

(β) Βρείτε την εντροπία της Y.

Απάντηση:

Η κατανομή της Y είναι η εξής:

$$Y = \begin{cases} -2 & \text{με πιθανότητα } \frac{1}{4} \\ 0 & \text{με πιθανότητα } \frac{1}{2} \\ +2 & \text{με πιθανότητα } \frac{1}{4} \end{cases}$$

Επομένως,

$$H(Y) = \sum_{y \in \mathcal{Y}} p(y) \log \frac{1}{p(y)} = 2 \times \frac{1}{4} \log 4 + \frac{1}{2} \log 2 = 1.5 \text{ bits}.$$

(γ) Για *οποιαδήποτε* κατανομή της N_2 (όχι, κατ' ανάγκη ομοιόμορφη όπως στο σχήμα), υποθέτοντας, πάντοτε, ότι η N_2 είναι τυχαία και ανεξάρτητη των X και N_1, δώστε το καλύτερο κάτω φράγμα που μπορείτε για την $H(Z)$. Θεωρούμε ότι η κατανομή της N_2 (και, επομένως, και η εντροπία της) είναι γνωστή. Προσοχή: Σε αυτό το ερώτημα (αντίθετα με όλα τα άλλα) δε θεωρούμε ότι η N_2 είναι, κατ' ανάγκη, δυαδική.

Απάντηση:

Η N_2 είναι ανεξάρτητη της Y δεδομένου ότι είναι ανεξάρτητη από τις X και N_1 και η Y είναι ντετερμινιστική συνάρτηση των και N_1. Στην 1η σειρά ασκήσεων είδαμε ότι όταν δύο ανεξάρτητες τ.μ. αθροίζονται, η εντροπία του αθροίσματος είναι τουλάχιστον ίση με την εντροπία κάθε μίας από τις μεταβλητές που αθροίζονται. Συνεπώς, $H(Z) \geq H(Y)$ και $H(Z) \geq H(N_2)$. Συνδυάζοντας τις δύο σχέσεις, $H(Z) \geq \max\{1.5, H(N_2)\}$.

(δ) Υπολογίστε, τώρα, την $H(Z)$ για $N_2 \sim \text{Unif}\{-1, +1\}$ και επαληθεύστε ότι υπερβαίνει το κάτω φράγμα που βρήκατε στο προηγούμενο ερώτημα.

Απάντηση:

Η ακολουθεί την κατανομή

$$Z = \begin{cases} -3 & \text{με πιθανότητα } \frac{1}{8} \\ -1 & \text{με πιθανότητα } \frac{3}{8} \\ +1 & \text{με πιθανότητα } \frac{3}{8} \\ +3 & \text{με πιθανότητα } \frac{1}{8} \end{cases}$$

Επομένως,

$$\begin{aligned} H(Z) &= \sum_{z \in \mathcal{Z}} p(z) \log \frac{1}{p(z)} \\ &= 2 \times \frac{1}{8} \log 8 + 2 \times \frac{3}{8} \log \frac{8}{3} = 3 - \frac{3}{4} \log 3 \approx 1.8113 \text{ bits}. \end{aligned}$$

Εναλλακτικά, από την αρχή διαχωρισιμότητας της εντροπίας, και παρατηρώντας ότι η πιθανότητα η Z να είναι αρνητική είναι ίση με την πιθανότητα να είναι θετική,

$$\begin{aligned} H(Z) &= H(S) + H(Z|S) = 1 + H(1/4) \\ &= 1 + \frac{1}{4} \log 4 + \frac{3}{4} \log \frac{4}{3} = 3 - \frac{3}{4} \log 3 \approx 1.8113 \text{ bits}, \end{aligned}$$

όπου S τ.μ. που δηλώνει το πρόσημο της Z.

Τέλος, μπορούμε να εφαρμόσουμε την αρχή διαχωρισιμότητας της εντροπίας και με διαφορετικό τρόπο, παρατηρώντας ότι, για δεδομένη απόλυτη τιμή της Z, η αρνητική και η θετική τιμή είναι ισοπίθανες (επιβεβαιώστε ως άσκηση).

(ε) Υπολογίστε τις $I(Y;X)$, $I(Z;X)$ και $I(Y,Z;X)$. Συγκρίνετέ τις μεταξύ τους και σχολιάστε.

Απάντηση:

• $I(Y;X) = H(Y) - H(Y|X) = H(Y) - H(X + N_1|X) = H(Y) - H(N_1) = 1.5 - 1 = 0.5$ bits. Όπως θα δούμε στα επόμενα μαθήματα, μπορούμε να δούμε το κανάλι μεταξύ X και Y ως ένα κανάλι διαγραφής. Όταν $Y = 0$, δεν μπορούμε να βρούμε ποια ήταν η τιμή της X. Ωστόσο, όταν $Y = -2$ γνωρίζουμε με βεβαιότητα ότι $X = -1$ (αντιστοίχως για $Y = +2$). Θα αποδείξουμε σε επόμενο μάθημα ότι η χωρητικότητα του καναλιού διαγραφής ισούται με $1 - \alpha$ και επιτυγχάνεται με ομοιόμορφη κατανομή της εισόδου X. Εδώ, $\alpha = \Pr\{Y = 0\} = 1/2$. Επομένως, $C = 0.5$ bits και επιτυγχάνουμε τη χωρητικότητα του καναλιού.

• $I(Z;X) = H(Z) - H(Z|X)$. Για δεδομένη $X = x$ μπορούμε να δούμε εύκολα ότι οι μάζες πιθανότητας $p_{Z|X}(z|x)$ είναι ίδιες με αυτές της κατανομής $p_Y(z)$ (αν και οι τιμές διαφέρουν). Επομένως, $H(Z|X) = H(Y)$. Εναλλακτικά, για δεδομένη $X = x$, το κανάλι από τη N_1 στη Z είναι το ίδιο με το κανάλι από μη δεδομένη X στην Y όπου το ρόλο της X έχει τώρα η N_1, και η γνωστή τιμή της X απλώς μεταθέτει την κατανομή της εξόδου κατά X. Ένας άλλος τρόπος είναι γράφοντας $H(Z|X) = H(Y + N_2|X) = H(X + N_1 + N_2|X) = H(N_1 + N_2) = 1.5$, γιατί η $N_1 + N_2$ ακολουθεί ίδια κατανομή με την Y. Συνεπώς, $I(Z;X) \approx 1.8113 - 1.5 = 0.3113$ bits.

• Από τον κανόνα αλυσίδας για την αμοιβαία πληροφορία, $I(Y,Z;X) = I(Y;X) + I(Z;X|Y)$. $I(Z;X|Y) = H(Z|Y) - H(Z|X,Y) = H(Z|Y) - H(Z|Y) = 0$. Επομένως, $I(Y,Z;X) = I(Y;X) = 0.5$ bits.

Συγκρίνοντας τις τιμές που βρήκαμε, παρατηρούμε ότι $I(Y;X) > I(Z;X)$. Αυτό είναι διαισθητικά λογικό δεδομένου ότι η Z περιέχει περισσότερο "θόρυβο" σε σχέση με τη X απ' ό,τι η Y. Μπορούμε, λοιπόν, να "μάθουμε" περισσότερα για τη X από την απ' ό,τι από τη Z (και αντίστροφα).

Παρατηρούμε, επίσης, ότι $I(Z;X|Y) = 0$. Δηλαδή, εάν γνωρίζουμε την Y, γνώση της X δε μας δίνει καμία περαιτέρω πληροφορία για τη Z. Και αυτό είναι διαισθητικά λογικό: $Z = Y + N_2$, και η N_2 είναι ανεξάρτητη της X. Επομένως, από τη στιγμή που γνωρίζουμε την Y, η μόνη αβεβαιότητα που απομένει είναι η τιμή της N_2 για την οποία η X δεν μπορεί να παράσχει καμία πληροφορία.

Τέλος, $I(Y,Z;X) = I(Y;X)$. Επομένως, όλη η πληροφορία που η X "περνάει" στα μετέπειτα στάδια περιέχεται ήδη στην Y. Παρατήρηση της Z επιπλέον της Y δε μας δίνει καμία παραπάνω πληροφορία για τη X.

Η άσκηση αποτελεί ένα παράδειγμα στο οποίο ισχύει η ανισότητα επεξεργασίας δεδομένων (data-processing inequality). Αποδεικνύεται ότι, εάν $X \to Y \to Z$ (σχηματίζουν, δηλαδή, αλυσίδα Markov), $I(Y;X) \geq I(Z;X)$. Μπορείτε να επαληθεύσετε ότι, στην άσκηση, $X \to Y \to Z$. Η ανισότητα επεξεργασίας δεδομένων έχει ενδιαφέρουσες προεκτάσεις, μια από τις οποίες είναι ότι δεν μπορούμε να δημιουργήσουμε νέα πληροφορία με ντετερμινιστική επεξεργασία τ.μ. Αντίθετα, στη γενική περίπτωση, "καταστρέφουμε" πληροφορία (ένα παράδειγμα είδατε στην 1η σειρά ασκήσεων στην περίπτωση μη αντιστρέψιμης συνάρτησης). Περισσότερα στο βιβλίο των Cover & Thomas, και στο μάθημα "Προχωρημένα Θέματα Θεωρίας Πληροφορίας".

(στ) Σας δίνονται οι εξής επιλογές: Παρατήρηση της τ.μ. Y μόνο, παρατήρηση της τ.μ. Z μόνο ή παρατήρηση του ζεύγους τ.μ. (Y, Z). Ποια παρατήρηση θα επιλέξετε ώστε να πάρετε όση περισσότερη πληροφορία για τη X μπορείτε (κατά μέσο όρο);

Απάντηση:

Από την παραπάνω συζήτηση προκύπτει ότι θα επιλέξουμε να παρατηρήσουμε την Y, δεδομένου ότι $I(Y;X) > I(Z;X)$ και του ότι ταυτόχρονη παρατήρηση της Z δε μας προσφέρει περισσότερη πληροφορία.

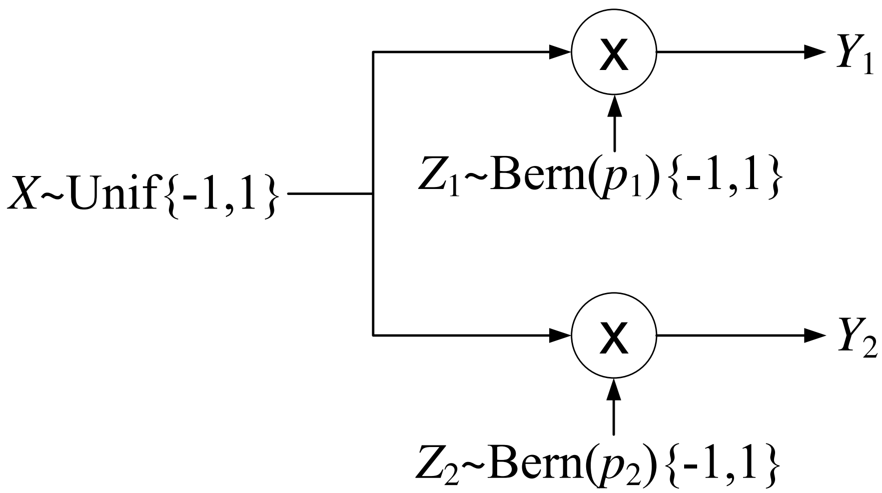
Παρατηρήσεις: Γενικά, οι επιδόσεις σε αυτήν την άσκηση ήταν καλές. Ένα λάθος που έκαναν οι περισσότεροι (μάλλον λόγω πίεσης χρόνου) ήταν στην κατανομή της Z. Η τιμή $Y = 0$ είναι πιο πιθανή από τις $Y = \pm 2$ και, επομένως, η κατανομή της Z είναι $\{-3, -1, +1, +3\} \leftrightarrow \{1/8, 3/8, 3/8, 1/8\}$ και όχι $\{1/6, 1/3, 1/3, 1/6\}$.

14. **Σύστημα με δύο εξόδους (Τελικό διαγώνισμα Θ. Π., Φεβρουάριος 2009)**

Θεωρούμε το σύστημα του Σχήματος 8. Η πηγή *δεν* έχει μνήμη και παίρνει τιμές $+1$ ή -1 με την ίδια πιθανότητα $1/2$. Οι *πολλαπλασιαστικοί* θόρυβοι Z_1 και Z_2 παίρνουν, επίσης, τιμές στο σύνολο $\{-1, 1\}$, αλλά ακολουθούν κατανομή Bernoulli(p_i). Δηλαδή,

$$Z_i = \begin{cases} +1 & \text{με πιθανότητα} \quad p_i \\ -1 & \text{με πιθανότητα} \quad 1 - p_i \end{cases} .$$

Οι τ.μ. Z_1 και Z_2 είναι ανεξάρτητες της X. Ωστόσο, στη γενική περίπτωση, **δεν** είναι ανεξάρτητες μεταξύ τους.



Σχήμα 8: Σύστημα με δύο εξόδους.

(α) Βρείτε την $H(X)$ και τις $H(Z_i)$ (συναρτήσει των p_i).

Απάντηση:

$H(X) = 1$ bit, δεδομένου ότι η X ακολουθεί ομοιόμορφη κατανομή.

Οι Z_i ακολουθούν κατανομή Bernoulli. Επομένως, $H(Z_i) = H(p_i) = -p_i \log p_i - (1 - p_i) \log(1 - p_i)$.

(β) Για *δεδομένα* p_1 και p_2, βρείτε ένα άνω φράγμα για την $H(Z_1, Z_2)$, καθώς και την από κοινού συνάρτηση μάζας πιθανότητας $p(z_1, z_2)$ η οποία επιτυγχάνει το άνω φράγμα.

Απάντηση:

Από τον κανόνα αλυσίδας για την εντροπία, $H(Z_1, Z_2) = H(Z_1) + H(Z_2|Z_1) \leq H(Z_1) + H(Z_2)$, με την ισότητα να ικανοποιείται όταν οι Z_1 και Z_2 είναι ανεξάρτητες. Συνεπώς, $H(Z_1, Z_2) \leq H(p_1) + H(p_2)$. Το άνω φράγμα επιτυγχάνεται από την από κοινού σ.μ.π. του Πίνακα 1.

	$Z_1 = +1$	$Z_1 = -1$
$Z_2 = +1$	$p_1 p_2$	$(1 - p_1)p_2$
$Z_2 = -1$	$p_1(1 - p_2)$	$(1 - p_1)(1 - p_2)$

Πίνακας 1: Από κοινού σ.μ.π. που επιτυγχάνει το άνω φράγμα για την $H(Z_1, Z_2)$.

(γ) Βρείτε τις $I(X; Y_1)$ και $I(X; Y_2)$. Για δεδομένα p_1 και p_2 υπάρχει τρόπος να αυξήσετε τις $I(X; Y_1)$ και $I(X; Y_2)$ αλλάζοντας την κατανομή της X; Επιτρέπεται να αλλάξετε μόνο τις πιθανότητες με τις οποίες $X = 1$ ή -1. Δεν μπορείτε να αυξήσετε το πλήθος τιμών, $|\mathcal{X}|$, της πηγής.

Απάντηση:

$$
\begin{aligned}
I(X; Y_i) &= H(Y_i) - H(Y_i|X) = H(Y_i) - H(X \cdot Z_i|X) \\
&\overset{(i)}{=} H(Y_i) - H(Z_i|X) \\
&\overset{(ii)}{=} H(Y_i) - H(Z_i) = H(Y_i) - H(p_i).
\end{aligned}
$$

(i) Ο πολλαπλασιασμός με μη μηδενική τιμή είναι αντιστρέψιμος, (ii) Οι Z_i είναι ανεξάρτητες της X. Με πράξεις μπορείτε εύκολα να δείτε ότι, για οποιαδήποτε τιμή της p_i, οι Y_1 και Y_2 ακολουθούν ομοιόμορφη κατανομή στο $\{-1, +1\}$. Ο πιο σύντομος τρόπος για να το διαπιστώσετε είναι με χρήση των αποτελεσμάτων για τα συμμετρικά κανάλια. Το κανάλι από τη X στην Y_1 είναι ένα δυαδικό συμμετρικό κανάλι (BSC). Επομένως, όταν στην είσοδο εφαρμόζεται ομοιόμορφη κατανομή η έξοδός του είναι, επίσης, ομοιόμορφα κατανεμημένη. Το ίδιο ισχύει και για το κανάλι από τη X στην Y_2. Συνεπώς,

$$I(X; Y_i) = 1 - H(p_i).$$

Δεδομένου ότι η χωρητικότητα του δυαδικού συμμετρικού καναλιού επιτυγχάνεται με ομοιόμορφη κατανομή για οποιαδήποτε τιμή της p_i, δεν υπάρχει τρόπος να αυξήσουμε τις $I(X; Y_1)$ και $I(X; Y_2)$ για δεδομένα p_1 και p_2. Το ίδιο αποτέλεσμα προκύπτει παρατηρώντας ότι η $H(Z_i|X)$ δεν εξαρτάται από τη X (λόγω της συμμετρίας) και η $H(Y_i)$ δεν μπορεί να υπερβεί το 1 bit (αυτό ακριβώς κάναμε όταν υπολογίσαμε τη χωρητικότητα του BSC).

(δ) Συγκρίνετε την $I(X; Y_1, Y_2)$ με τις $I(X; Y_i)$. Είναι μικρότερη; Μεγαλύτερη; Ίση; Ποια είναι η μεγαλύτερη τιμή (σε bits) που μπορεί να πάρει η $I(X; Y_1, Y_2)$;

Υπόδειξη: Στο ερώτημα αυτό μπορείτε να απαντήσετε χωρίς να βρείτε την τιμή της $I(X;Y_1,Y_2)$.

Απάντηση:

Από τον κανόνα αλυσίδας για την αμοιβαία πληροφορία,

$$I(X;Y_1,Y_2) = I(Y_1,Y_2;X) = I(Y_1;X) + I(Y_2;X|Y_1).$$

Γνωρίζουμε, επίσης, ότι η αμοιβαία πληροφορία (και η δεσμευμένη αμοιβαία πληροφορία) είναι πάντοτε μη αρνητική. Επομένως,

$$I(Y_2;X|Y_1) \geq 0 \Rightarrow I(X;Y_1,Y_2) \geq I(X;Y_1).$$

Για τους ίδιους λόγους,

$$I(X;Y_1,Y_2) \geq I(X;Y_2).$$

Επομένως, στη γενική περίπτωση, όπως περιμένουμε και διαισθητικά, παίρνουμε περισσότερη πληροφορία αν κοιτάξουμε και τις 2 εξόδους. Αυτό οφείλεται στο ότι, στη γενική περίπτωση, οι τιμές των θορύβων Z_1 και Z_2 δεν είναι πλήρως εξαρτημένες μεταξύ τους.

Υπάρχουν περιπτώσεις στις οποίες δεν ωφελεί να κοιτάξουμε και τις δύο εξόδους. Η μία είναι όταν οι Z_1 και Z_2 είναι πλήρως εξαρτημένες, δηλαδή $Z_1 = Z_2$ (πάντοτε) ή $Z_1 = -Z_2$ (πάντοτε). Στην περίπτωση αυτή $I(Y_1;X|Y_2) = H(Y_1|Y_2) - H(Y_1|Y_2,X) = H(X+Z_1|X+Z_2) - H(X+Z_1|X+Z_2,X) = 0 - 0 = 0$ και, επομένως, $I(X;Y_1,Y_2) = I(X;Y_1)$. Ομοίως, αποδεικνύεται ότι $I(X;Y_1,Y_2) = I(X;Y_2)$. Η άλλη περίπτωση είναι η τετριμμένη, όπου ο θόρυβος είναι ντετερμινιστικός, δηλαδή κάποιο από τα p_i ισούται με 0 ή 1. Κοιτάζοντάς το και πάλι μαθηματικά, εάν υποθέσουμε ότι η Z_1 είναι ντετερμινιστική, $I(Y_2;X|Y_1) = H(Y_2|Y_1) - H(Y_2|Y_1,X) = 0$, γιατί, από τη στιγμή που γνωρίζουμε την Y_1 μπορούμε να υπολογίσουμε και τη X αφαιρώντας τη γνωστή Z_1 και, επομένως, $H(Y_2|Y_1) = H(Y_2|Y_1,X)$.

Η μεγαλύτερη τιμή που μπορεί να πάρει η $I(X;Y_1,Y_2)$ είναι 1 bit.

$$I(X;Y_1,Y_2) = H(X) - H(X|Y_1,Y_2) \leq H(X) = 1 \text{ bit}.$$

Το αποτέλεσμα είναι λογικό: Η πληροφορία που "μπαίνει" στο σύστημα ισούται ακριβώς με 1 bit. Επομένως, η μέγιστη πληροφορία στην έξοδο η οποία προέρχεται από τη X (<u>όχι</u> η συνολική πληροφορία που περιέχεται στην έξοδο) δεν μπορεί να υπερβαίνει το 1 bit. Ωστόσο, στη γενική περίπτωση, πρέπει να "συλλέξουμε" την πληροφορία από όλες τις εξόδους του συστήματος.

Η ισότητα ισχύει όταν μπορούμε να βρούμε τη X από τις Y_1 και Y_2 χωρίς αβεβαιότητα. Αυτό συμβαίνει όταν τουλάχιστον μια από τις Z_1 και Z_2 είναι ντετερμινιστική ή όταν $Z_1 = -Z_2$ (οπότε $Y_1 + Y_2 = 2X$).

(ε) Υποθέστε, τώρα, ότι οι Z_1 και Z_2 είναι *ανεξάρτητες* μεταξύ τους. Βρείτε μια έκφραση για την $(X;Y_1,Y_2)$ συναρτήσει των p_1 και p_2.

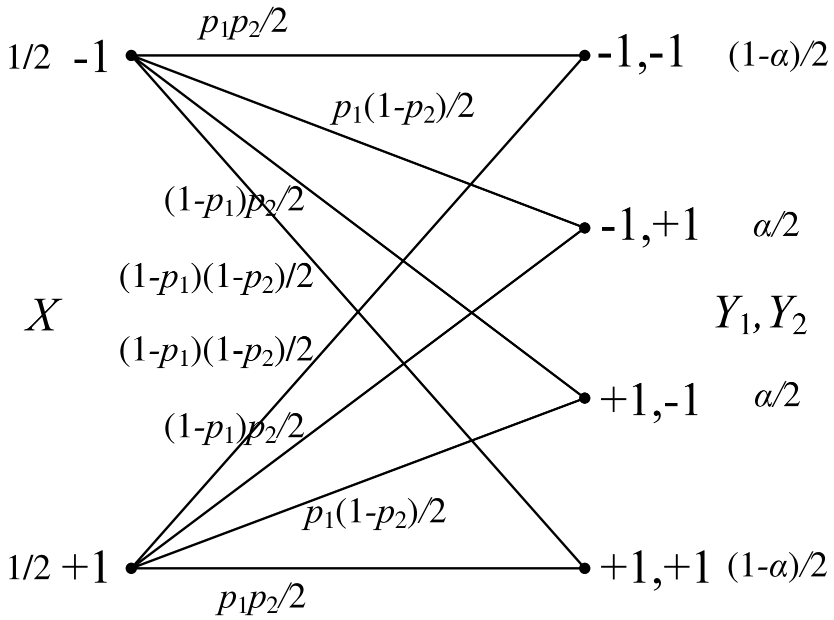
Υπόδειξη: Ενδέχεται να σας φανεί χρήσιμη η αντικατάσταση $\alpha \triangleq p_1 + p_2 - 2p_1p_2$.

Απάντηση:

Γνωρίζουμε ότι

$$I(X; Y_1, Y_2) = H(Y_1, Y_2) - H(Y_1, Y_2|X).$$

Προκειμένου να υπολογίσουμε την $H(Y_1, Y_2)$ βρίσκουμε την από κοινού κατανομή, $p(y_1, y_2)$, των Y_1 και Y_2.



Σχήμα 9: Υπολογισμός $p(y_1, y_2)$ και $p(y_1, y_2|x)$.

Με χρήση του θεωρήματος ολικής πιθανότητας (Σχήμα 9) και θέτοντας $\alpha \triangleq p_1 + p_2 - 2p_1p_2$,

$$\begin{aligned}
p(-1,-1) = p(+1,+1) &= \frac{1}{2}\left(1 - p_1 - p_2 + 2p_1p_2\right) = \frac{(1-\alpha)}{2}, \\
p(-1,+1) = p(+1,-1) &= \frac{1}{2}\left(p_1 + p_2 - 2p_1p_2\right) = \frac{\alpha}{2}.
\end{aligned}$$

Συνεπώς,

$$\begin{aligned}
H(Y_1, Y_2) &= -2 \times \frac{\alpha}{2} \log \frac{\alpha}{2} - 2 \times \frac{(1-\alpha)}{2} \log \frac{(1-\alpha)}{2} \\
&= -\alpha \log \alpha - (1-\alpha)\log(1-\alpha) + \alpha + (1-\alpha) \\
&= 1 + H(\alpha).
\end{aligned}$$

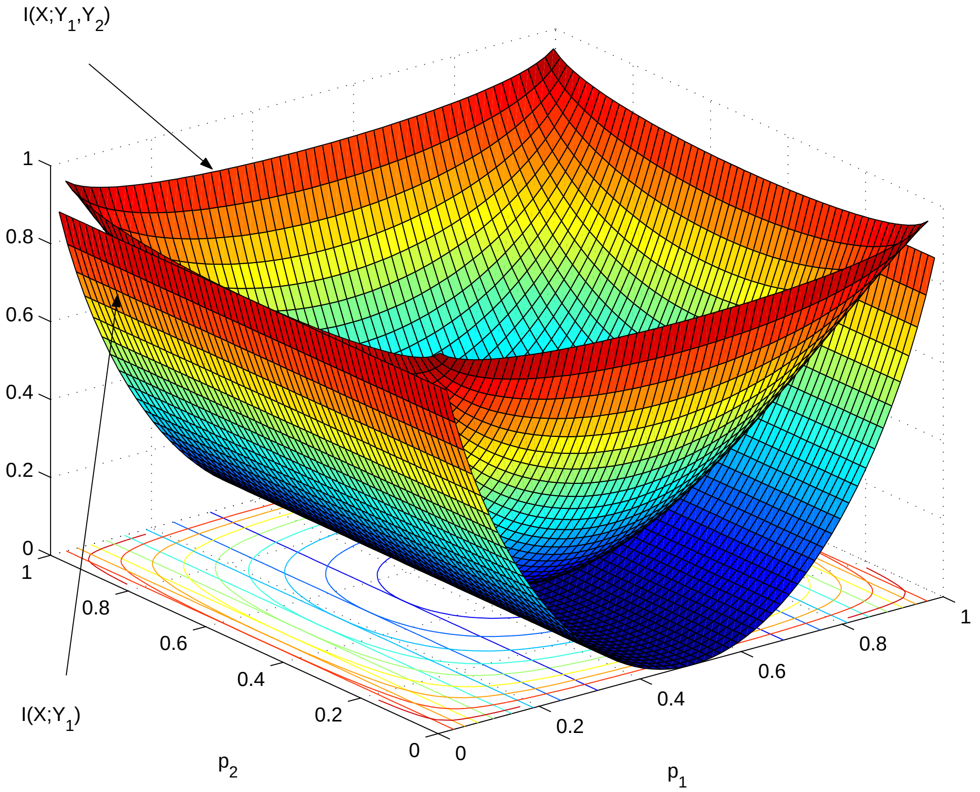
Με χρήση του κανόνα αλυσίδας,

$$\begin{aligned}
H(Y_1, Y_2|X) = H(Y_1|X) + H(Y_2|X, Y_1) &= H(X \cdot Z_1|X) + H(X \cdot Z_2|X, X \cdot Z_1) = \\
&= H(Z_1) + H(Z_2|Z_1) \overset{(i)}{=} H(Z_1) + H(Z_2) \\
&= H(p_1) + H(p_2).
\end{aligned}$$

(i) από την υπόθεση ότι οι Z_1 και Z_2 είναι ανεξάρτητες.
Αντικαθιστώντας,

$$I(X; Y_1, Y_2) = 1 + H(\alpha) - H(p_1) - H(p_2).$$

Η $I(X; Y_1, Y_2)$ συναρτήσει των p_1 και p_2 έχει σχεδιαστεί στο Σχήμα 10. Παρατηρήστε ότι $I(X; Y_1, Y_2) = 0$ όταν $p_1 = p_2 = \frac{1}{2}$. Παρατηρήστε, επίσης, ότι $I(X; Y_1, Y_2) \geq I(X; Y_1)$, όπως αποδείχτηκε στο προηγούμενο ερώτημα.



Σχήμα 10: $I(X; Y_1, Y_2)$ για ανεξάρτητες Z_1 και Z_2 και σύγκριση με την $I(X; Y_1)$.

Παρατηρήσεις: Η άσκηση αυτή δυσκόλεψε αρκετούς και ήταν μάλλον η δυσκολότερη του διαγωνίσματος. Ένα λάθος που έκαναν κάποιοι είναι να θέσουν συγκεκριμένες τιμές για τα p_i στο Ερώτημα (β). Ωστόσο, σύμφωνα με την εκφώνηση και όπως διευκρινίστηκε και κατά τη διάρκεια της εξέτασης, οι τιμές των p_i είναι *δεδομένες*. Το γεγονός ότι οι τιμές των p_i (δηλαδή οι περιθώριες της $p(z_1, z_2)$) είναι δεδομένες δε σημαίνει ότι η $p(z_1, z_2)$ είναι μοναδική. Επίσης, ένας από τους σκοπούς της άσκησης ήταν να συνειδητοποιήσετε ότι, στη γενική περίπτωση, οι Y_1 και Y_2 **δεν** είναι ανεξάρτητες ακόμα και εάν οι θόρυβοι είναι ανεξάρτητοι. Η τιμή και των δύο εξαρτάται από τη X. Εάν, για παράδειγμα, $p_1 = p_2 = 0.001$, γνώση της τιμής της Y_1 μας βοηθά να εκτιμήσουμε την Y_2 με μικρή πιθανότητα σφάλματος. Η μόνη περίπτωση που οι Y_1 και Y_2 είναι ανεξάρτητες είναι όταν $p_1 = p_2 = 1/2$ **και** οι Z_i είναι ανεξάρτητες. Τέλος, προσοχή: $H(X \cdot Z_1|X) = H(Z_1)$, γιατί αν μας αποκαλυφθεί η απομένει η αβεβαιότητα της Z_1

(αρκεί το X να μην είναι 0). Αντίθετα, $H(X|X\cdot Z_1)\neq H(X|Z_1)$. Και πάλι, αν $Z_1=+1$ με πολύ μεγάλη πιθανότητα, παίρνουμε σχεδόν όλη την πληροφορία για τη X από το γινόμενο $X\cdot Z_1$ (κατά μέσο όρο), ενώ, αντίθετα, δεν παίρνουμε καμία πληροφορία για τη X μόνο από τη Z_1.

15. **Ανισότητες (Επαναληπτική Εξέταση Θ. Π., Σεπτέμβριος 2009)**

(α) Να αποδειχτεί ότι, για $n\geq 2$,

$$H(X_1,X_2,\ldots,X_n)\geq \sum_{i=1}^{n} H(X_i|X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n).$$

Στη συνέχεια, να αποδειχτεί ότι

$$\frac{1}{2}\left[H(X_1,X_2)+H(X_2,X_3)+H(X_1,X_3)\right]\geq H(X_1,X_2,X_3).$$

Υπόδειξη: Χρησιμοποιήστε τις σχέσεις

$$H(X_1,X_2,X_3)=H(X_i,X_j)+H(X_k|X_i,X_j),$$

καθώς και την πρώτη σχέση που αποδείξατε.

Απάντηση:

Από τον κανόνα αλυσίδας για την εντροπία,

$$\begin{aligned}
H(X_1,X_2,\ldots,X_n) &= H(X_1)+H(X_2|X_1)+\ldots+H(X_n|X_1,X_2,\ldots,X_{n-1})\\
&\overset{(a)}{\geq} H(X_1|X_2,X_3,\ldots,X_n)+H(X_2|X_1,X_3,\ldots,X_n)+\ldots\\
&\quad + H(X_n|X_1,X_2,\ldots,X_{n-1})\\
&= \sum_{i=1}^{n} H(X_i|X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n).
\end{aligned}$$

(a) $H(X|Y)\leq H(X)$.

Εναλλακτική απόδειξη με χρήση επαγωγής (από συνάδελφό σας):

• Για $n=2$, $H(X_1,X_2)=H(X_1)+H(X_2|X_1)\geq H(X_1|X_2)+H(X_2|X_1)$.

• Έστω ότι ισχύει για $n=k$:

$H(X_1,X_2,\ldots,X_k)\geq \sum_{i=1}^{k} H(X_i|X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_k)$.

• Για $n=k+1$,

$$\begin{aligned}
H(X_1,X_2,\ldots,X_{k+1}) &= H(X_1,X_2,\ldots,X_k)+H(X_{k+1}|X_1,X_2,\ldots,X_k)\\
&\overset{(b)}{\geq} \sum_{i=1}^{k} H(X_i|X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_k)\\
&\quad + H(X_{k+1}|X_1,X_2,\ldots,X_k)\\
&= \sum_{i=1}^{k+1} H(X_i|X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_k).
\end{aligned}$$

(b) από την υπόθεση για $n = k$.

Αθροίζοντας, τώρα, τις σχέσεις $H(X_1, X_2, X_3) = H(X_i, X_j) + H(X_k|X_i, X_j)$ για όλους τους συνδυασμούς τριών τ.μ.,

$$\begin{aligned}
3H(X_1, X_2, X_3) &= H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3) \\
&\quad + \sum_{i=1}^{3} H(X_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \\
&\overset{(c)}{\leq} H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3) + H(X_1, X_2, X_3) \\
\Rightarrow 2H(X_1, X_2, X_3) &\leq H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3) \\
\Rightarrow H(X_1, X_2, X_3) &\leq \frac{1}{2}\left[H(X_1, X_2) + H(X_2, X_3) + H(X_1, X_3)\right].
\end{aligned}$$

(c) Από την πρώτη σχέση που αποδείχτηκε.

(β) Έστω $p_{XY}(x, y)$ και $q_{XY}(x, y)$ δύο κατανομές μάζας πιθανότητας στο σύνολο $\mathcal{X} \times \mathcal{Y}$. Να αποδειχτεί ότι $D(p_{XY}||q_{XY}) \geq D(p_X||q_X)$, όπου $p_X(x)$ και $q_X(x)$ οι περιθώριες σ.μ.π. των $p_{XY}(x, y)$ και $q_{XY}(x, y)$, αντιστοίχως.

Απάντηση (Cover & Thomas Theorem 2.5.3):

Από τον ορισμό της σχετικής εντροπίας,

$$\begin{aligned}
D(p_{XY}(x, y)||q_{XY}(x, y)) &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{q_{XY}(x, y)} \\
&= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p_{XY}(x, y) \log \frac{p_X(x)p_{Y|X}(y|x)}{q_X(x)q_{Y|X}(y|x)} \\
&= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p_{XY}(x, y) \log \frac{p_X(x)}{q_X(x)} + \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p_{XY}(x, y) \log \frac{p_{Y|X}(y|x)}{q_{Y|X}(y|x)} \\
&= \sum_{x\in\mathcal{X}} \log \frac{p_X(x)}{q_X(x)} \sum_{y\in\mathcal{Y}} p_{XY}(x, y) + D(p_{Y|X}||q_{Y|X}) \\
&= \sum_{x\in\mathcal{X}} p_X(x) \log \frac{p_X(x)}{q_X(x)} + D(p_{Y|X}||q_{Y|X}) \\
&= D(p_X||q_X) + D(p_{Y|X}||q_{Y|X}) \geq D(p_X||q_X),
\end{aligned}$$

δεδομένου ότι η σχετική εντροπία είναι πάντοτε μη αρνητική.

Εναλλακτική απόδειξη (από συνάδελφό σας):

$$\begin{aligned}
D(p_X(x)||q_X(x)) &= \sum_{x\in\mathcal{X}} p_X(x) \log \frac{p_X(x)}{q_X(x)} \\
&= \sum_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} p_{X,Y}(x, y)\right) \log \frac{\sum_{y\in\mathcal{Y}} p_{X,Y}(x, y)}{\sum_{y\in\mathcal{Y}} q_{X,Y}(x, y)} \\
&\overset{(d)}{\leq} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p_{X,Y}(x, y) \log \frac{p_{X,Y}(x, y)}{q_{X,Y}(x, y)} = D(p_{XY}(x, y)||q_{XY}(x, y)).
\end{aligned}$$

(d) από την ανισότητα log-sum (Cover & Thomas Theorem 2.7.1).

16. **Ισότητες και ανισότητες (Επαναληπτική εξέταση Π. Θ. Θ. Π., Σεπτέμβριος 2009)**

Συμπληρώστε τις παρακάτω σχέσεις με $=$, $\leq$ ή $\geq$. Δικαιολογήστε την απάντησή σας. Στην περίπτωση που ισχύει $\leq$ ή $\geq$, προσδιορίστε πότε ισχύει η ισότητα.

Σημείωση: Απαντήσεις που δεν είναι επαρκώς αιτιολογημένες δε βαθμολογούνται.

(α) $I(X;Y)$? $I(g(X);Y)$. Η $g()$ είναι ντετερμινιστική συνάρτηση.

Απάντηση:

$$I(X;Y) = H(Y) - H(Y|X) \geq H(Y) - H(Y|g(X)) = I(g(X);Y),$$

δεδομένου ότι η $g(X)$ δεν μπορεί να περιέχει περισσότερη πληροφορία για την Y απ' ό,τι η X.

Ένας άλλος τρόπος επίλυσης (στην ουσία ο ίδιος) είναι να παρατηρήσετε ότι $Y \to X \to g(X)$ και να χρησιμοποιήσετε την ανισότητα επεξεργασίας δεδομένων.

Η ισότητα ισχύει όταν η $g(X)$ είναι 1-προς-1, δηλαδή όταν από την τιμή της $g(X)$ μπορούμε να προσδιορίσουμε την τιμή της X.

(β) $I(Y;Z|X)$? $I(Y;Z)$ εάν $p(x,y,z) = p(x)p(y)p(z|x,y)$.

Απάντηση:

$$\begin{aligned} I(Y;Z|X) &= H(Y|X) - H(Y|X,Z) \stackrel{(i)}{=} H(Y) - H(Y|X,Z) \\ &\geq H(Y) - H(Y|Z) = I(Y;Z). \end{aligned}$$

(i) Οι X και Y είναι ανεξάρτητες.

Η ισότητα ισχύει όταν $H(Y|X,Z) = H(Y|Z) \Rightarrow H(Y|Z) - H(Y|X,Z) = 0 \Rightarrow I(X;Y|Z) = 0$, όταν, δηλαδή, οι X και Y είναι υπό συνθήκη ανεξάρτητες δεδομένης της Z.

(γ) $H(X|Z)$? $H(X|Y) + H(Y|Z)$.

Απάντηση:

$$H(X|Y) + H(Y|Z) \geq H(X|Y,Z) + H(Y|Z) = H(X,Y|Z) \geq H(X|Z).$$

Η ισότητα ισχύει όταν $H(X|Y) = H(X|Y,Z)$ και $H(X,Y|Z) = H(X|Z)$, όταν, δηλαδή $I(X;Z|Y) = 0$ και $H(Y|X,Z) = 0$.

17. **Μεταβολή εντροπίας (Επαναληπτική Εξέταση Π. Θ. Θ. Π., Σεπτέμβριος 2010)**

Θεωρήστε μια κατανομή $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ με $N \geq 2$ ενδεχόμενα, όλα *μη μηδενικής πιθανότητας*, δηλαδή $p_n > 0\, \forall n$. Θα δείξουμε ότι, για οποιαδήποτε κατανομή, υπάρχει πάντοτε τουλάχιστον ένα ενδεχόμενο, έστω K, το οποίο αν αφαιρέσουμε θα ελαττώσουμε την εντροπία. Δηλαδή, υπάρχει πάντοτε τουλάχιστον ένα p_k έτσι ώστε, αν $\mathbf{p}' = \left(\frac{p_1}{1-p_K}, \frac{p_2}{1-p_K}, \ldots, \frac{p_{K-1}}{1-p_K}, \frac{p_{K+1}}{1-p_K}, \ldots, \frac{p_N}{1-p_K}\right)$, $H(\mathbf{p}') < H(\mathbf{p})$.

(α) Δείξτε ότι δεν μπορούμε πάντοτε να ελαττώσουμε την εντροπία αν επιλέξουμε το ενδεχόμενο τυχαία. Δηλαδή, υπάρχουν κατανομές για τις οποίες υπάρχει p τέτοιο ώστε $H(\mathbf{p}') > H(\mathbf{p})$.

Απάντηση:

Αρκεί ένα αντιπαράδειγμα. Έστω, για παράδειγμα, η κατανομή $\left((1-p), \frac{p}{N-1}, \frac{p}{N-1}, \ldots, \frac{p}{N-1}\right)$ από την οποία αφαιρούμε το πρώτο ενδεχόμενο. Από τη διαχωρισιμότητα της εντροπίας, $H(\mathbf{p}) = H(p) + pH\left(\frac{p}{N-1}, \frac{p}{N-1}, \ldots, \frac{p}{N-1}\right) = H(p) + p\log(N-1)$. Θέλουμε $H(\mathbf{p}') > H(\mathbf{p}) \Rightarrow H(p) < (1-p)\log(N-1) \Rightarrow H(p)/(1-p) > \log(N-1)$. Για $p = 1/2$, αρκεί $\log_2(N-1) > 2 \Rightarrow N > 5$. Πράγματι, για $p = 2$ και $N = 5$, $H(\mathbf{p}) \approx 2.161$ bits $< 2.322 \approx H(\mathbf{p}')$.

(β) Στη συνέχεια θα αποδείξετε ένα λήμμα που θα σας βοηθήσει να αποδείξετε το ζητούμενο. Δείξτε ότι, για $p \in (0, 1/2]$, η συνάρτηση $\frac{H(p)}{p}$, όπου $H(p) \triangleq -p\log_2 p - (1-p)\log_2(1-p)$ είναι φθίνουσα συνάρτηση του p.

Απάντηση:

Παραγωγίζοντας,

$$
\begin{aligned}
\frac{H(p)}{p} &= -\log_2 p - \frac{1-p}{p}\log_2(1-p) \Rightarrow \\
\frac{\partial}{\partial p}\frac{H(p)}{p} &= -\frac{1}{p}\log_2 e + \frac{1}{p^2}\log_2(1-p) + \frac{1-p}{p}\frac{1}{1-p}\log_2 e \\
&= \frac{\log_2(1-p)}{p^2} < 0 \text{ για } p \in (0, 1/2].
\end{aligned}
$$

Επειδή η παράγωγος είναι πάντοτε αρνητική στο διάστημα $(0, 1/2]$, η $H(p)/p$ είναι φθίνουσα.

(γ) Εξηγήστε γιατί, σε μία οποιαδήποτε κατανομή με ενδεχόμενα, $(p_1, p_2, \ldots, p_N)$, όλα *μη μηδενικά*, υπάρχει πάντοτε τουλάχιστον ένα ενδεχόμενο, L, με $p_L \leq \frac{1}{N}$.

Απάντηση:

Έστω $p_{\min} \triangleq \min\{p_n\}$ η ελάχιστη μάζα της κατανομής. Δεδομένου ότι $\sum_{n=1}^{N} p_n = 1 \Rightarrow \sum_{n=1}^{n} p_{\min} \leq 1 \Rightarrow Np_{\min} \leq 1 \Rightarrow p_{\min} \leq \frac{1}{N}$.

(*δ) Χρησιμοποιώντας τα Ερωτήματα (β) και (γ) (ή κάποιον άλλο τρόπο, αν προτιμάτε) αποδείξτε ότι υπάρχει πάντοτε ένα ενδεχόμενο, K, το οποίο αν αφαιρεθεί η εντροπία της κατανομής μειώνεται.

Απάντηση:

Θέλουμε να δείξουμε ότι $H(\mathbf{p}') < H(\mathbf{p})$. Έστω, για απλοποίηση των εξισώσεων, ότι αναδιατάσσουμε τα ενδεχόμενα και το ενδεχόμενο που αφαιρούμε είναι το

N-στό. Από τον ορισμό της εντροπίας,

$$\begin{aligned}
H(\mathbf{p}') &= -\sum_{n=1}^{N-1} \frac{p_n}{1-p_N} \log \frac{p_n}{1-p_N} = -\sum_{n=1}^{N-1} \frac{p_n}{1-p_N} \log p_n + \sum_{n=1}^{N-1} \frac{p_n}{1-p_N} \log(1-p_N) \\
&= -\sum_{n=1}^{N} \frac{p_n}{1-p_N} \log p_n + \frac{p_N}{1-p_N} \log p_N + \log(1-p_N) \\
&= \frac{1}{1-p_N} H(\mathbf{p}) + \frac{1}{1-p_N} \left[p_N \log p_N + (1-p_N)\log(1-p_N)\right] \\
&= \frac{1}{1-p_N} \left[H(\mathbf{p}) - H(p_N)\right].
\end{aligned}$$

Επομένως, αρκεί να δείξουμε ότι υπάρχει τουλάχιστον ένα p_N ώστε

$$\begin{aligned}
&H(\mathbf{p}') < H(\mathbf{p}) \Rightarrow \frac{1}{1-p_N} \left[H(\mathbf{p}) - H(p_N)\right] < (\mathbf{p}) \Rightarrow \\
&H(p_N) > p_N H(\mathbf{p}) \Rightarrow H(\mathbf{p}) < \frac{H(p_N)}{p_N}.
\end{aligned}$$

Όπως δείξαμε στο Ερώτημα (β), η $\frac{H(p_N)}{p_N}$ είναι φθίνουσα. Επίσης, από το Ερώτημα (γ), υπάρχει $p_N \leq \frac{1}{N}$. Συνεπώς, υπάρχει p_N τέτοιο ώστε

$$\begin{aligned}
\frac{H(p_N)}{p_N} &\geq N \cdot H(1/N) = N \left(\frac{1}{N} \log N + \frac{N-1}{N} \log \frac{N}{N-1} \right) \\
&= \log N + (N-1)\log N - (N-1)\log(N-1) \\
&= N \log N - (N-1)\log(N-1) \\
&> N \log N - (N-1)\log N = \log N \geq H(\mathbf{p}).
\end{aligned}$$

18. **Ανισότητες (Τελική Εξέταση Θ. Π., Φεβρουάριος 2011)**

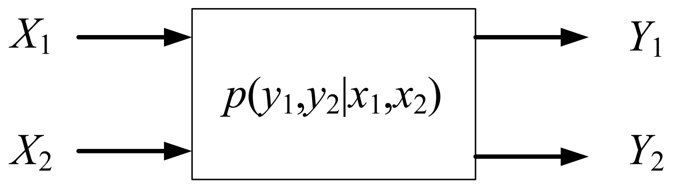
    Θεωρήστε 4 διακριτές τ.μ. X_1, X_2, Y_1 και Y_2 οι οποίες ικανοποιούν τις εξής συνθήκες

    - Οι X_1 και X_2 είναι *ανεξάρτητες*. Δηλαδή, $p_{X_1 X_2}(x_1, x_2) = p_{X_1}(x_1) \cdot p_{X_2}(x_2)$.
    - Οι Y_1 και Y_2 εξαρτώνται από τις X_1 και X_2 μέσω της $p_{Y_1 Y_2|X_1 X_2}(y_1, y_2|x_1, x_2)$.

    Για παράδειγμα, οι Y_1 και Y_2 μπορεί να είναι έξοδοι του συστήματος του Σχήματος 11, στο οποίο εισάγουμε ανεξάρτητες εισόδους (δηλαδή ο χρήστης 1 δεν μπορεί να συνεννοηθεί με το χρήστη 2). Ωστόσο, στη γενική περίπτωση, η έξοδος Y_j εξαρτάται όχι μόνο από τη X_j, αλλά και από τη X_i, $i \neq j$.

    (α) Δείξτε ότι $I(X_1, X_2; Z) \geq I(X_1; Z) + I(X_2; Z)$, όπου $Z = Y_1$, $Z = Y_2$ ή $Z = (Y_1, Y_2)$.

    Δηλαδή, κάποιος που διαθέτει πρόσβαση στα X_1 και X_2 ταυτοχρόνως, μπορεί να αντλήσει περισσότερη πληροφορία για το Z από τη συνολική πληροφορία που μπορούν να αντλήσουν δύο άτομα που διαθέτουν πρόσβαση ο καθένας μόνο στο X_1 και μόνο στο X_2.



Σχήμα 11: Στοχαστικό σύστημα με εισόδους X_1 και X_2 και εξόδους Y_1 και Y_2.

Πότε ίσχύει η ισότητα στην παραπάνω σχέση;

Απάντηση:

Από τον κανόνα αλυσίδας για την αμοιβαία πληροφορία,

$$\begin{aligned}
I(X_1, X_2; Z) &= I(X_1; Z) + I(X_2; Z|X_1) \\
&= I(X_1; Z) + H(X_2|X_1) - H(X_2|X_1, Z) \\
&\overset{(i)}{=} I(X_1; Z) + H(X_2) - H(X_2|X_1, Z) \\
&\overset{(ii)}{\geq} I(X_1; Z) + H(X_2) - H(X_2|Z) \\
&= I(X_1; Z) + I(X_2; Z).
\end{aligned}$$

(i) Οι X_1 και X_2 είναι ανεξάρτητες. (ii) Conditioning cannot increase the entropy.
Η ισότητα στην παραπάνω σχέση ισχύει όταν
$H(X_2|X_1, Z) = H(X_2|Z) \Rightarrow H(X_2|Z) - H(X_2|X_1, Z) = I(X_1; X_2|Z) = 0.$
Γνωρίζουμε ότι αυτό ισχύει εάν και μόνο εάν οι X_1 και X_2 είναι ανεξάρτητες *δεδομένης* της Z.

(β) Έστω, τώρα, ότι ισχύει $p(y_1, y_2|x_1, x_2) = p(y_1|x_1, x_2) \cdot p(y_2|x_1, x_2)$. Δηλαδή, οι Y_1 και Y_2 είναι ανεξάρτητες *δεδομένων των* X_1 και X_2.
Δείξτε ότι

$$I(X_1, X_2; Y_1, Y_2) \leq I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2).$$

Δηλαδή, κάποιο μέρος της πληροφορίας που περιέχει η Y_1 για τις X_1 και X_2 υπάρχει και στην Y_2 (και αντιστρόφως).
Πότε ισχύει η ισότητα;

Απάντηση:

Χρησιμοποιώντας, και πάλι, τον κανόνα αλυσίδας,

$$\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2|Y_1) \\
&= I(X_1, X_2; Y_1) + H(Y_2|Y_1) - H(Y_2|Y_1, X_1, X_2) \\
&\overset{(i)}{\leq} I(X_1, X_2; Y_1) + H(Y_2) - H(Y_2|Y_1, X_1, X_2) \\
&\overset{(ii)}{=} I(X_1, X_2; Y_1) + H(Y_2) - H(Y_2|X_1, X_2) \\
&= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2).
\end{aligned}$$

(i) Conditioning cannot increase the entropy, (ii) Οι Y_1 και Y_2 είναι ανεξάρτητες *δεδομένων των* X_1 και X_2.
Η ισότητα ισχύει όταν $H(Y_2|Y_1) = H(Y_2)$, όταν, δηλαδή, οι Y_1 και Y_2 είναι ανεξάρτητες (και όχι απλώς ανεξάρτητες δεδομένων των X_1 και X_2).

$\boxed{\text{Παρατηρήσεις:}}$

Κάποιοι δεν πρόσεξαν ότι στο Ερώτημα (α), η ισότητα ισχύει όταν οι X_1 και X_2 είναι ανεξάρτητες *δεδομένης* της Z. Η ανεξαρτησία των X_1 και X_2 (η οποία, άλλωστε, είχε υποτεθεί στο πρόβλημα) δεν αρκεί.

19. **Σύγκριση εντροπιών (Τελική Εξέταση Θ. Π., Φεβρουάριος 2012)**

Θεωρούμε *δυαδική* πηγή *χωρίς* μνήμη, η οποία παράγει τα συμβολα 0 και 1 με πιθανότητες p και $1-p$, αντιστοίχως. Έστω ότι συμβολίζουμε την έξοδο της πηγής με X_i.

Για κάθε ζεύγος εξόδων $X_{2k}X_{2k+1}$ της πηγής, ορίζουμε την τ.μ. C_k η οποία ισούται με τον αριθμό των '0' στο ζεύγος $X_{2k}X_{2k+1}$. Δηλαδή, οι πιθανές τιμές της C_k είναι 0, 1 και 2.

(α) Ποια είναι η κατανομή της C_k;
Αν θεωρήσουμε ότι οι C_k είναι έξοδοι μίας πηγής (η οποία παράγει σύμβολα με το μισό ρυθμό σε σχέση με την πηγή που παράγει τις X_i), η πηγή έχει μνήμη;
$\boxed{\text{Απάντηση:}}$
Οι C_k είναι ανεξάρτητες επειδή οι X_i είναι ανεξάρτητες. Επομένως, η νέα πηγή δεν έχει μνήμη και για όλες τις C_k (ανεξαρτήτως k) ισχύει

$$p_C(c) = \begin{cases} (1-p)^2 & \text{όταν} \quad c=0 \\ 2p(1-p) & \text{όταν} \quad c=1 \\ p^2 & \text{όταν} \quad c=2 \end{cases}$$

(β) Βρείτε την εντροπία του *ζεύγους* $(X_{2k}X_{2k+1})$, την εντροπία της C_k και τη διαφορά τους. Σχολιάστε.
Υπόδειξη: Η άσκηση λύνεται γρήγορα αν εφαρμόσετε την αρχή διαχωρισιμότητας της εντροπίας.
$\boxed{\text{Απάντηση:}}$

$$\begin{aligned} H(X_{2k}, X_{2k+1}, C_k) &= H(X_{2k}, X_{2k+1}) + H(C_k|X_{2k}, X_{2k+1}) \\ &= H(C_k) + H(X_{2k}, X_{2k+1}|C_k) \\ \overset{(a)}{\Rightarrow} H(X_{2k}, X_{2k+1}) &= H(C_k) + H(X_{2k}, X_{2k+1}|C_k) \\ \Rightarrow H(C_k) &= H(X_{2k}, X_{2k+1}) - H(X_{2k}, X_{2k+1}|C_k) \end{aligned}$$

(a) Αν γνωρίζουμε τις X_{2k} και X_{2k+1} μπορούμε να υπολογίσουμε τη C_k.
Επειδή η πηγή που παράγει τις X_i δεν έχει μνήμη και είναι δυαδική, $H(X_{2k}, X_{2k+1}) = 2H(p)$ bits.
Για να υπολογίσουμε την $H(X_{2k}, X_{2k+1}|C_k)$ παρατηρούμε ότι, όταν $C_k = 0$ ή $C_k = 2$ γνωρίζουμε ακριβώς τις X_{2k} και X_{2k+1}. Αλλιώς χρειαζόμαστε πληροφορία 1 bit για να μάθουμε αν $X_{2k} = 0$ και $X_{2k+1} = 1$ ή το ανάποδο. Συνεπώς,

$$H(X_{2k}, X_{2k+1}|C_k) = \mathrm{Pr}\{C_k = 1\}H(X_{2k}, X_{2k+1}|C_k = 1) = 2p(1-p).$$

Παρατηρούμε ότι, όπως περιμέναμε, η εντροπία δεν μπορεί να αυξηθεί εάν εφαρμόσουμε κάποια (ντετερμινιστική) συνάρτηση σε μία ή περισσότερες τ.μ. Στη συγκεκριμένη, μάλιστα, περίπτωση, η εντροπία ελαττώνεται γιατι επεξεργαζόμαστε τις X_{2k} και X_{2k+1} με μη αντιστρέψιμο τρόπο.

(γ) Για ποια τιμή του p μεγιστοποιείται η διαφορά $H(X_{2k}, X_{2k+1}) - H(C_k)$; Υπάρχουν τιμές του p για τις οποίες $H(X_{2k}, X_{2k+1}) = H(C_k)$;

Απάντηση:

Η διαφορά $H(X_{2k}, X_{2k+1}) - H(C_k)$ μεγιστοποιείται όταν $p = \frac{1}{2}$. Αυτό έχει ως αποτέλεσμα τα ζεύγη 01 και 10 να εμφανίζονται με (συνολική) πιθανότητα $\frac{1}{2}$.

Στην τετριμμένη περίπτωση όπου $p = 0$ ή $p = 1$, $H(X_{2k}, X_{2k+1}) = H(C_k) = 0$

(δ) Επαναλάβετε το Ερώτημα (β) για την περίπτωση τριάδων εξόδων της πηγής. Δηλαδή, ορίζουμε τ.μ. D_k που ισούται με τον αριθμό των '0' στην τριάδα $(X_{3k}, X_{3k+1}, X_{3k+2})$.

Απάντηση:

Εφαρμόζουμε και πάλι την αρχή διαχωρισιμότητας της εντροπίας. Για να υπολογίσουμε την $H(X_{3k}, X_{3k+1}, X_{3k+2}|D_k)$ παρατηρούμε ότι, αν $D_k = 1$ χρειαζόμαστε την πληροφορία της θέσης του 0 ($\log_2 3$ bits), ενώ αν $D_k = 2$, χρειαζόμαστε την πληροφορία της θέσης του 1 (και πάλι $\log_2 3$ bits). Επομένως, $H(X_{3k}, X_{3k+1}, X_{3k+2}) = 3$ bits, $H(X_{3k}, X_{3k+1}, X_{3k+2}|D_k) = (3p(1-p)^2 + 3p^2(1-p)) \log_2 3$ bits και $H(D_k) = 3 - (3p(1-p)^2 + 3p^2(1-p)) \log_2 3$ bits.

Με πράξεις, $H(X_{3k}, X_{3k+1}, X_{3k+2}|D_k) = (3p(1-p)^2 + 3p^2(1-p)) \log_3 = 3p(1-p) \log 3$.

(ε) Γενικεύστε για την περίπτωση όπου η τ.μ. E_k ορίζεται ως ο αριθμός των '0' στη n-άδα $(X_{nk}, X_{nk+1}, \ldots, X_{nk+n-1})$. Αρκεί να δώσετε μία (σωστή) γενική σχέση.

Υπόδειξη: $(x+y)^n = x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{2}x^2y^{n-2} + \binom{n}{1}xy^{n-1} + y^n$.

Απάντηση:

• Με πιθανότητα $\binom{n}{1}p^{n-1}(1-p) = np^{n-1}(1-p)$, $E_k = n-1$. Σε αυτήν την περίπτωση, χρειαζόμαστε $\log_2 n$ bits για να μάθουμε ποιο από τα n (ισοπίθανα) ενδεχόμενα εμφανίστηκε. Το ίδιο και στην περίπτωση όπου $E_k = 1$.

• Με πιθανότητα $\binom{n}{2}p^{n-2}(1-p)^2$, $E_k = n-2$. Σε αυτήν την περίπτωση, χρειαζόμαστε $\log_2 \binom{n}{2}$ bits για να μάθουμε ποιο από τα $\binom{n}{2}$ (ισοπίθανα) ενδεχόμενα εμφανίστηκε. Το ίδιο και στην περίπτωση όπου $E_k = 2$.

• Συνεπώς, για $E_k = n-l$ ή $E_k = l$, χρειαζόμαστε $\log_2 \binom{n}{l}$ bits.

Επομένως,

$$H(X_{nk}, X_{nk+1}, \ldots, X_{nk+n-1}|E_k) = \sum_{l=1}^{n-1} \binom{n}{l} p^l (1-p)^{n-l} \log_2 \binom{n}{l}.$$

20. **Φράγματα για την εντροπία (Επαναληπτική Εξέταση Θ. Π., Σεπτέμβριος 2012)**

Θεωρούμε διακριτή τ.μ. X με συνάρτηση μάζας πιθανότητας (pmf) $\mathbf{p} = \{p_1, p_2, \ldots, p_N\}$ και $p_1 \leq p_2 \leq \cdots \leq p_N$.

(α) Δείξτε ότι $H(X) \geq \sum_{n=1}^{N} p_n(1-p_n)$ nats $\geq 1 - p_N$ nats.

Υπόδειξη: Χρησιμοποιήστε την ανισότητα $\ln x \leq x - 1$, $x > 0$.

$\boxed{\text{Απάντηση:}}$

$$
\begin{aligned}
H(X) &= \sum_{n=1}^{N} -p_n \ln p_n \\
&\overset{(a)}{\geq} \sum_{n=1}^{N} p_n(1-p_n) \\
&\overset{(b)}{\geq} \sum_{n=1}^{N} p_n(1-p_N) = 1-p_N.
\end{aligned}
$$

(a) $\ln p_n \leq p_n - 1 \Rightarrow -\ln p_n \geq 1 - p_n$,
(b) $p_N \geq p_n \, \forall n \Rightarrow 1 - p_N \leq 1 - p_n \, \forall n$.

(β) Δείξτε ότι $H(X) \geq H(p_N)$, όπου $H(p_N) \triangleq p_N \log \frac{1}{p_N} + (1-p_N) \log \frac{1}{1-p_N}$.

Πότε ισχύει η ισότητα;

$\boxed{\text{Απάντηση:}}$

Ένας τρόπος είναι με χρήση της αρχής διαχωρισιμότητας της εντροπίας. Έστω τ.μ. Y τέτοια ώστε

$$
Y = \begin{cases} 0 & \text{εάν} \quad X \neq N \\ 1 & \text{εάν} \quad X = N \end{cases}
$$

$H(X,Y) = H(Y) + H(X|Y) = H(X) + H(Y|X) = H(X)$.
Επομένως,

$$
H(X) = H(Y) + H(X|Y) \overset{(a)}{\geq} H(Y) = H(p_N).
$$

(a) $H(X|Y) \geq 0$.
Η ισότητα ισχύει όταν $H(X|Y) = 0$, όταν, δηλαδή, όλα τα p_n εκτός από το p_N και ένα ακόμα (έστω το p_1) ισούνται με 0.

(γ) Δείξτε ότι $H(X) \geq -\log p_N$.
Πότε ισχύει η ισότητα;

$\boxed{\text{Απάντηση:}}$

$$
H(X) = \sum_{n=1}^{N} p_n \log \frac{1}{p_n} \overset{(a)}{\geq} \sum_{n=1}^{N} p_n \log \frac{1}{p_N} = \log \frac{1}{p_N}.
$$

(a) $p_N \geq p_n \, \forall n \Rightarrow \frac{1}{p_N} \leq \frac{1}{p_n} \, \forall n$.
Η ισότητα ισχύει όταν όλα τα ενδεχόμενα είναι ισοπίθανα: $p_1 = p_2 = \cdots = p_N = \frac{1}{N}$.

(δ) **Πιο χρονοβόρο**
Δείξτε ότι $H(X) \geq 2(1-p_N)$ bits.
Υπόδειξη: Χρησιμοποιήστε το φράγμα του Ερωτήματος (β) για την περίπτωση όπου $p_N \geq \frac{1}{2}$ και το φράγμα του Ερωτήματος (γ) για την περίπτωση όπου $p_N \leq \frac{1}{2}$.

$\boxed{\text{Απάντηση:}}$

Από το Ερώτημα (β), υποθέτοντας ότι $p_N \geq \frac{1}{2}$,

$$\begin{aligned} H(X) \geq H(p_N) &= p_N \log \frac{1}{p_N} + (1-p_N) \log \frac{1}{1-p_N} \\ &\stackrel{(a)}{\geq} (1-p_n)\left(\log \frac{1}{p_N} + \log \frac{1}{1-p_N}\right) \\ &= (1-p_N) \log \frac{1}{p_N(1-p_N)}. \end{aligned}$$

(a) $p_N \geq \frac{1}{2} \Rightarrow 1 - p_N \leq \frac{1}{2} \leq p_N$.

Αρκεί να δείξουμε ότι $\log \frac{1}{p_N(1-p_N)} \geq 2 \Leftrightarrow p_N(1-p_N) \leq \frac{1}{4}$.

Αλλά $4p_N^2 - 4p_N + 1 = (2p_N - 1)^2$. Επομένως, $4p_N^2 - 4p_N + 1 \geq 0 \Rightarrow p_N(1-p_N) \leq \frac{1}{4}$ για όλα τα p_N (η ισότητα ισχύει για $p_N = 1/2$.

Αν, τώρα, $p_N \leq \frac{1}{2}$, από το Ερώτημα (γ),

$$H(X) \geq \log \frac{1}{p_N}.$$

Αρκεί να δείξουμε ότι $\frac{1}{p_N} \geq 2(1-p_N)$ για $p_N \leq \frac{1}{2}$. Η παράγωγος της συνάρτησης $\log_2 \frac{1}{p_N} - 2 + 2p_N = -\log_2 p_N - 2 + 2p_N$ είναι αρνητική για $p_N < \log_2 e/2$ και, επομένως, αρνητική για $p_N \leq \frac{1}{2}$. Επομένως, η συνάρτηση $\log_2 \frac{1}{p_N} - 2 + 2p_N$ είναι φθίνουσα για $p_N \leq \frac{1}{2}$. Επίσης, για $p_N = \frac{1}{2}$, $\log_2 \frac{1}{p_N} = 1 = 2(1-p_N)$. Συνεπώς, $\log_2 \frac{1}{p_N} - 2 + 2p_N \geq 0$ για $p_N \leq \frac{1}{2}$.

21. **Μεταβολή εντροπίας (Επαναληπτική Εξέταση Θ. Π., Σεπτέμβριος 2013)**

Θεωρούμε διακριτή κατανομή με μάζες $p_1, p_2, \ldots, p_N$. Έστω ότι η κατανομή δεν είναι ομοιόμορφη, δηλαδή υπάρχουν τουλάχιστον δύο μάζες που δεν είναι ίσες.

Χωρίς βλάβη της γενικότητας, έστω ότι οι μάζες αυτές είναι οι p_1 και p_2. Επίσης, υποθέτουμε ότι $p_1 > p_2$.

Έστω, τώρα, ότι μεταβάλλουμε κατά μία μικρή ποσότητα $\Delta p > 0$ τις p_i ως εξής:

$$\begin{aligned} p_1' &= p_1 - \Delta p \\ p_2' &= p_2 + \Delta p \\ p_i' &= p_i,\ i \neq 1, 2. \end{aligned}$$

Επίσης, θεωρούμε ότι $p_1 - \Delta p \geq p_2 + \Delta p$.

Δηλαδή, κινούμαστε προς την κατεύθυνση της εξίσωσης των p_1 και p_2 κρατώντας τις άλλες μάζες σταθερές.

Δείξτε ότι οποιαδήποτε τέτοια απόπειρα εξίσωσης μαζών κατανομής (δηλαδή οποιαδήποτε κίνηση προς την ομοιόμορφη κατανομή) αυξάνει την εντροπία της κατανομής. Δηλαδή, δείξτε ότι

$$H(p_1', p_2', \ldots, p_N') > H(p_1, p_2, \ldots, p_N).$$

Απάντηση:

Από τον ορισμό της εντροπίας,

$$\begin{aligned}
H(p'_1, p'_2, \ldots, p'_N) &= -\sum_{i=1}^{n} p'_i \log p'_i \\
&= -(p_1 - \Delta p)\log(p_1 - \Delta p) - (p_2 + \Delta p)\log(p_2 + \Delta p) - \sum_{i=3}^{n} p_i \log p_i.
\end{aligned}$$

Επομένως, αν θεωρήσουμε ότι το Δp είναι αρκετά μικρό και παραγωγίσουμε,

$$\frac{\partial}{\partial \Delta p} H(p'_1, p'_2, \ldots, p'_N) = \log(p_1 - \Delta p) - \log(p_2 + \Delta p) = \log \frac{p_1 - \Delta p}{p_2 + \Delta p} \geq 0,$$

δεδομένου ότι $p_1 - \Delta p \geq p_2 + \Delta p$.

Επομένως, ξεκινώντας από την $H(p_1, p_2, \ldots, p_N)$ και μεταβάλλοντας τις p_1 και p_2 κατά μικρό βήμα Δp, η παράγωγος παραμένει θετική (δηλαδή η εντροπία αυξάνει) μέχρι να εξισωθούν οι p_1 και p_2.

22. **Σύντομες Ερωτήσεις (Τελική Εξέταση Θ. Π., Φεβρουάριος 2014)**

(α) Έστω X διακριτή τ.μ. Συγκρίνετε την $H(X)$ με την $H(X+3)$.

Απάντηση:

Δεδομένου ότι η εντροπία διακριτής τ.μ. εξαρτάται μόνο από την κατανομή και όχι από τις τιμές της, $H(X) = H(X+3)$. Εναλλακτικά, αν $f(x) = x+3$, επειδή η $f(x)$ είναι αντιστρέψιμη, $H(f(X)) = H(X)$.

(β) Έστω ότι κωδικοποιούμε *από κοινού* n σύμβολα πηγής $X_1, X_2, \ldots, X_n$ με χρήση κώδικα Huffman.

Αν $c(X_1, X_2, \ldots, X_n)$ είναι η κωδική λέξη που αντιστοιχεί στην ακολουθία $X_1, X_2, \ldots, X_n$, συγκρίνετε την $H(X_1, X_2, \ldots, X_n)$ με την $H\left(c(X_1, X_2, \ldots, X_n)\right)$.

Απάντηση:

Δεδομένου ότι ο κώδικας Huffman είναι μοναδικώς αποκωδικοποιήσιμος, η συνάρτηση κωδικοποίησης $c(\cdot)$ είναι αντιστρέψιμη. Επομένως, $H(X_1, X_2, \ldots, X_n) = H\left(c(X_1, X_2, \ldots, X_n)\right)$.

(γ) Στο προηγούμενο ερώτημα, έστω, τώρα, ότι $l(X_1, X_2, \ldots, X_n)$ είναι η συνάρτηση μήκους, δηλαδή το μήκος της κωδικής λέξης $c(X_1, X_2, \ldots, X_n)$. Συγκρίνετε την $H(X_1, X_2, \ldots, X_n)$ με την $H\left(l(X_1, X_2, \ldots, X_n)\right)$.

Απάντηση:

Στη γενική περίπτωση, ενδέχεται να έχουν το ίδιο μήκος περισσότερες από μία διαφορετικές κωδικές λέξεις. Επομένως, η $l(\cdot)$ δεν είναι, απαραιτήτως, αντιστρέψιμη και $H(X_1, X_2, \ldots, X_n) \geq H\left(l(X_1, X_2, \ldots, X_n)\right)$.

(δ) Έστω $\{X_n\}$ *στάσιμη* στοχαστική διαδικασία με $X_n \in \mathcal{X}$ και $H(X_n) < \infty$. Διατάξτε τις $H(X_1)$, $H(X_2|X_1)$ και $\frac{1}{2}H(X_1, X_2)$. Δηλαδή αντιστοιχίστε τις 3 ποσότητες στα a, b και c ώστε $a \flat b \sharp c$, όπου τα $\flat$ και $\sharp$ αντιστοιχούν σε $\leq$ ή $<$.

Απάντηση:

Από τη στασιμότητα, $H(X_2) = H(X_1)$. Συνεπώς, $H(X_2|X_1) \leq H(X_1)$.
Επίσης,

$$\frac{1}{2}H(X_1, X_2) = \frac{1}{2}H(X_1) + \frac{1}{2}H(X_2|X_1) \leq \frac{1}{2}H(X_1) + \frac{1}{2}H(X_1) = H(X_1)$$

και

$$\frac{1}{2}H(X_1, X_2) = \frac{1}{2}H(X_1)+\frac{1}{2}H(X_2|X_1) \geq \frac{1}{2}H(X_2|X_1)+\frac{1}{2}H(X_2|X_1) = H(X_2|X_1).$$

Συνεπώς,

$$H(X_2|X_1) \leq \frac{1}{2}H(X_1, X_2) \leq H(X_1).$$

Οι ισότητες ισχύουν όταν η στοχαστική διαδικασία είναι i.i.d.

23. **► Ο σκύλος που ψάχνει για το κόκκαλο – Cover & Thomas 4.12**

Ένας σκύλος κινείται επάνω στους ακέραιους αριθμούς. Σε κάθε βήμα συνεχίζει προς την ίδια κατεύθυνση με πιθανότητα 0.9, ενώ αλλάζει κατεύθυνση με πιθανότητα 0.1. Υποθέστε ότι ο σκύλος ξεκινά από τη θέση $X_0 = 0$ και ότι το πρώτο βήμα μπορεί να γίνει προς οποιαδήποτε κατεύθυνση (θετική ή αρνητική) με την ίδια πιθανότητα. Για παράδειγμα, μια πιθανή διαδρομή του σκύλου είναι

$(X_0, X_1, \ldots) = (0, -1, -2, -3, -4, -3, -2, -1, 0, 1, 2, 3, 4, 3, ...)$.

(α) Υπολογίστε την $H(X_0, X_1, \ldots, X_n)$.

Απάντηση:

Από τον κανόνα αλυσίδας,

$$\begin{aligned} H(X_0, X_1, \ldots, X_n) &= \sum_{i=0}^{n} H(X_i|X_{i-1}, \ldots, X_0) \\ &= H(X_0) + H(X_1|X_0) + \sum_{i=2}^{n} H(X_i|X_{i-1}, \ldots, X_0). \end{aligned}$$

- $H(X_0) = 0$ δεδομένου ότι η θέση του σκύλου τη χρονική στιγμή 0 μας είναι γνωστή ($X_0 = 0$).
- $H(X_1|X_0) = H(1/2) = 1$, καθώς έχουμε υποθέσει ότι το πρώτο βήμα του σκύλου γίνεται προς μια από τις 2 κατευθύνσεις (-1 ή 1) με την ίδια πιθανότητα.
- Παρατηρούμε ότι, για τα επόμενα βήματα, η θέση του σκύλου τη χρονική στιγμή i εξαρτάται μόνο από τη θέση του τις χρονικές στιγμές $i-1$ και $i-2$, δεδομένου ότι, πριν προχωρήσει στο βήμα i, ο σκύλος είτε συνεχίζει στην ίδια κατεύθυνση με αυτή που είχε μεταξύ των βημάτων $i-2$ και $i-1$ ή αντιστρέφει την πορεία του. Σε κάθε περίπτωση, ο σκύλος λαμβάνει υπόψη του μόνο την κατεύθυνση που είχε μεταξύ των βημάτων $i-2$ και $i-1$, ανεξαρτήτως

της διαδρομής του πριν από το βήμα $i-2$. Μαθηματικά, η πορεία του σκύλου αποτελεί τυχαία διαδικασία Markov 2ης τάξης, και

$$\begin{aligned} p(x_i|x_{i-1},\ldots,x_0) &= p(x_i|x_{i-1},x_{i-2}) \\ &= \begin{cases} 0.9 & \text{όταν} \quad (x_i-x_{i-1})=(x_{i-1}-x_{i-2}),\ |x_i-x_{i-1}|=1 \\ 0.1 & \text{όταν} \quad (x_i-x_{i-1})=-(x_{i-1}-x_{i-2}),\ |x_i-x_{i-1}|=1 \\ 0 & \text{αλλιώς} \end{cases} \end{aligned}$$

Επομένως, $H(X_i|X_{i-1},X_{i-2},\ldots,X_0)=H(X_i|X_{i-1},X_{i-2})=H(0.9,0.1)=-0.9\log 0.9-0.1\log 0.1$ bits.

Αντικαθιστώντας στον κανόνα αλυσίδας,

$$\begin{aligned} H(X_0,X_1,\ldots,X_n) &= \sum_{i=0}^{n} H(X_i|X_{i-1},\ldots,X_0) \\ &= H(X_0)+H(X_1|X_0)+\sum_{i=2}^{n} H(X_i|X_{i-1},X_{i-2}) \\ &= 0+1+(n-1)H(0.9,0.1)=1+(n-1)H(0.9,0.1). \end{aligned}$$

(β) Υπολογίστε το ρυθμό εντροπίας, $H(\mathcal{X})$, της θέσης του σκύλου.

Απάντηση:

Από τον ορισμό του ρυθμού εντροπίας,

$$\begin{aligned} H(\mathcal{X}) &= \lim_{n\to\infty} \frac{1}{n+1} H(X_0,X_1,\ldots,X_n) \\ &= \lim_{n\to\infty} \left(\frac{1}{n+1}+\frac{n-1}{n+1}H(0.9,0.1)\right)=H(0.9,0.1). \end{aligned}$$

Παρατηρήστε ότι

$$\begin{aligned} H(\mathcal{X}) &= H(\mathcal{X}') \\ &= \lim_{n\to\infty} (X_n|X_{n-1},X_{n-2},\ldots,X_0) \\ &= \lim_{n\to\infty} (X_n|X_{n-1},X_{n-2}) = H(0.9,0.1). \end{aligned}$$

Το αποτέλεσμα είναι λογικό διαισθητικά, αφού σε κάθε βήμα αυτό που δεν ξέρουμε (η αβεβαιότητα) είναι ποια κατεύθυνση θα επιλέξει ο σκύλος μεταξύ 2 πιθανών κατευθύνσεων.

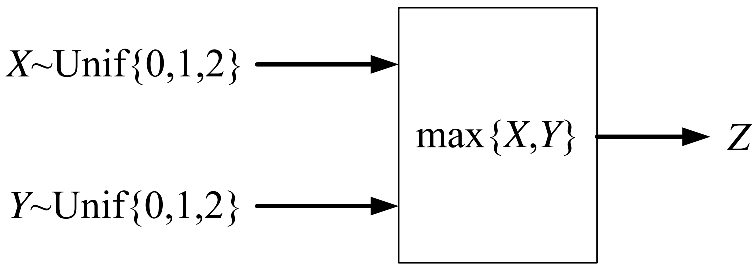
Παρατηρήστε, επίσης, ότι εάν δε λαμβάναμε υπόψη μας το γεγονός ότι ο σκύλος αποφασίζει μόνο με βάση τις 2 τελευταίες θέσεις του, θα χρειαζόμασταν περισσότερα bits για να κωδικοποιήσουμε τις κινήσεις του.

24. **Μέγιστο δύο τ.μ. (Πρόοδος Θ. Π., Νοέμβριος 2009)**

Θεωρούμε το ντετερμινιστικό σύστημα του Σχήματος 12. Οι τ.μ. X και Y ακολουθούν ομοιόμορφη κατανομή στο σύνολο $\mathcal{X}=\mathcal{Y}=\{0,1,2\}$. Οι X και Y είναι *ανεξάρτητες* μεταξύ τους. Το κύκλωμα υπολογίζει την έξοδο με χρήση της σχέσης $Z=\max\{X,Y\}$.

Δίνεται ότι $\log_2 3\approx 1.585$ και $\log_2 5\approx 2.3219$.

Εάν προτιμάτε, μπορείτε να αφήσετε στις απαντήσεις σας όρους της μορφής $H(\mathbf{p})=H(p_1,p_2,\ldots,p_K)$ χωρίς να τους υπολογίσετε, αρκεί να δώσετε τις τιμές των $p_1,p_2,\ldots,p_K$.



Σχήμα 12: Υπολογισμός μεγίστου δύο τ.μ.

(α) Υπολογίστε την $H(X)$ και την $I(X;Y)$.

Απάντηση:

Δεδομένου ότι η X ακολουθεί ομοιόμορφη κατανομή, $H(X) = \log_2 3$ bits.

Επειδή οι X και είναι ανεξάρτητες, $I(X;Y) = 0$.

(β) Υπολογίστε την $H(Z)$ και την $I(X;Z)$.

Απάντηση:

Για να βρούμε την $H(Z)$ βρίσκουμε, κατ' αρχάς, την κατανομή της Z. Παρατηρούμε ότι $Z = 0$ όταν $(X,Y) = (0,0)$, $Z = 1$ όταν $(X,Y) \in \{(0,1),(1,0),(1,1)\}$ και $Z = 2$ όταν $(X,Y) \in \{(0,2),(2,0),(1,2),(2,1),(2,2)\}$. Δεδομένου ότι όλα τα ζεύγη (X,Y) είναι ισοπίθανα, $\{p_Z(0), p_Z(1), p_Z(2)\} = \left\{\frac{1}{9}, \frac{3}{9}, \frac{5}{9}\right\}$. Εναλλακτικά, $\Pr\{Z = 0\} = \Pr\{X = 0 \text{ και } Y = 0\} = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$, $\Pr\{Z = 2\} = \Pr\{X = 2 \text{ ή } Y = 2\} = \Pr\{X = 2\} + \Pr\{Y = 2\} - \Pr\{X = 2 \text{ και } Y = 2\} = \frac{1}{3} + \frac{1}{3} - \frac{1}{9} = \frac{5}{9}$ και $\Pr\{Z = 1\} = 1 - \Pr\{Z = 0\} - \Pr\{Z = 2\} = \frac{3}{9}$. Επομένως,

$$
\begin{aligned}
H(Z) &= \frac{1}{9}\log 9 + \frac{1}{3}\log 3 + \frac{5}{9}\log\frac{9}{5} \\
&= \frac{2}{9}\log 3 + \frac{1}{3}\log 3 + \frac{10}{9}\log 3 - \frac{5}{9}\log 5 \\
&= \frac{15}{9}\log 3 - \frac{5}{9}\log 5 \approx 1.3516 \text{ bits.}
\end{aligned}
$$

Εναλλακτικά, από την αρχή διαχωρισιμότητας της εντροπίας,

$$
\begin{aligned}
H(Z) &= H(Z \overset{?}{=} 0) + p_Z(Z > 0) H(Z \overset{?}{=} 1 | Z \neq 0) \\
&= H\left(\frac{1}{9}\right) + \frac{8}{9} H\left(\frac{3}{8}\right) \\
&= \frac{1}{9}\log 9 + \frac{8}{9}\log\frac{9}{8} + \frac{8}{9}\left(\frac{3}{8}\log\frac{8}{3} + \frac{5}{8}\log\frac{8}{5}\right) \\
&= \frac{2}{9}\log 3 + \frac{16}{9}\log 3 - \frac{24}{9} + \frac{24}{9} - \frac{1}{3}\log 3 - \frac{5}{9}\log 5 \\
&= \frac{15}{9}\log 3 - \frac{5}{9}\log 5 \approx 1.3516 \text{ bits.}
\end{aligned}
$$

Στη συνέχεια υπολογίζουμε την $H(Z|X)$ που απαιτείται για τον υπολογισμό της

$I(X;Z)$

$$\begin{aligned}
H(Z|X) &= p_X(0)H(Z|X=0) + p_X(1)H(Z|X=1) + p_X(2)H(Z|X=2) \\
&= \frac{1}{3}H(Y) + \frac{1}{3}H(Y \leq 1) + \frac{1}{3} \times 0 \\
&= \frac{1}{3}\log 3 + \frac{1}{3}H\left(\frac{2}{3}\right) \\
&= \frac{1}{3}\log 3 + \frac{1}{3}\left(\frac{1}{3}\log 3 + \frac{2}{3}\log\frac{3}{2}\right) \\
&= \frac{1}{3}\log 3 + \frac{1}{3}\left(\log 3 - \frac{2}{3}\right) = \frac{2}{3}\log 3 - \frac{2}{9} \approx 0.8344 \text{ bits.}
\end{aligned}$$

Επομένως,

$$\begin{aligned}
I(X;Z) &= H(Z) - H(Z|X) \\
&= \frac{15}{9}\log 3 - \frac{5}{9}\log 5 - \frac{2}{3}\log 3 + \frac{2}{9} \\
&= \log 3 - \frac{5}{9}\log 5 + \frac{2}{9} \approx 0.5172 \text{ bits.}
\end{aligned}$$

(γ) Βρείτε ένα βέλτιστο *άμεσο δυαδικό* κώδικα για τη Z και συγκρίνετε το μέσο μήκος του με την $H(Z)$.

Απάντηση:

Ένας τρόπος να βρούμε το βέλτιστο κώδικα είναι με τη χρήση κωδικοποίησης Huffman. Ένας κώδικας Huffman είναι ο $\{0,1,2\} \leftrightarrow \{00, 01, 1\}$.

Το μέσο μήκος του κώδικα ισούται με $\mathbb{E}[l] = 2 \times \left(\frac{1}{9} + \frac{1}{3}\right) + \frac{5}{9} = \frac{13}{9} \approx 1.4444$ bits. Παρατηρούμε ότι επιτυγχάνεται συμπίεση με 0.0924 bits μακριά από την εντροπία, $H(Z)$.

(δ) Δείξτε ότι $I(X;Y|Z) > 0$ (*όχι* $\geq$). Συγκρίνετε με την $I(X;Y)$ του Ερωτήματος (α).

Υπόδειξη: Η απάντηση προκύπτει πολύ πιο σύντομα αν χρησιμοποιήσετε ιδιότητες εντροπίας ή/και αμοιβαίας πληροφορίας και δείτε τι ισχύει στο συγκεκριμένο πρόβλημα, παρά αν κάνετε πράξεις.

Απάντηση:

Η $I(X;Y|Z)$ μπορεί να γραφτεί στη μορφή $I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$. Επίσης, από τον κανόνα αλυσίδας για την εντροπία,

$$H(X,Y|Z) = H(Y|Z) + H(X|Y,Z) \Rightarrow -H(X|Y,Z) = H(Y|Z) - H(X,Y|Z).$$

Αντικαθιστώντας την πρώτη σχέση, προκύπτει ότι

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

Η παραπάνω σχέση είναι ανάλογη της (2.45) του βιβλίου των Cover & Thomas για δεσμευμένη αμοιβαία πληροφορία.

Γνωρίζουμε ότι $H(X,Y|Z) \leq H(X|Z) + H(Y|Z)$ με $=$ όταν οι X και Y είναι ανεξάρτητες *δεδομένης της* Z. Ωστόσο, αυτό δεν ισχύει. Για παράδειγμα, αν $Z =$

0, τότε γνωρίζουμε ότι η μόνη τιμή που επιτρέπεται για το ζεύγος (X, Y) είναι η (0,0). Συνεπώς, $H(X, Y|Z) < H(X|Z) + H(Y|Z)$ και $I(X; Y|Z) > 0$. Επομένως, αν γνωρίζουμε την τιμή της Z, η "αποκάλυψη" της X μας δίνει πληροφορία για την Y και αντιστρόφως (κατά μέσο όρο).

Η (απευθείας) απάντηση ότι $I(X; Y|Z) > 0$ όταν οι X και Y δεν είναι υπό συνθήκη ανεξάρτητες δεδομένης της Z γίνεται δεκτή, αρκεί να δείξετε ότι δεν ισχύει η υπό συνθήκη ανεξαρτησία.

(ε) Εάν τόσο ο συμπιεστής όσο και ο αποσυμπιεστής γνωρίζουν, με κάποιο τρόπο, την τιμή της Y (αλλά όχι της X), προτείνετε ένα βέλτιστο *άμεσο δυαδικό* κώδικα για τη συμπίεση της Z. Πόσο αποδοτικός είναι ο κώδικας;

Απάντηση:

Εδώ πρέπει να δείτε ότι πρέπει να συμπιέσουμε με χρήση της κατανομής $p_{Z|Y}$ και ότι το θεωρητικό κάτω όριο είναι η $H(Z|Y)$. Εάν $Y = 2$, $Z = 2$ με πιθανότητα 1, οπότε δε χρειάζεται να κωδικοποιήσουμε τη Z. Αν $Y = 0$, τότε η Z ακολουθεί ομοιόμορφη κατανομή με τιμές $\{0, 1, 2\}$. Ένας βέλτιστος κώδικας είναι ο $\{0, 10, 11\}$. Τέλος, εάν $Y = 1$, η Z ακολουθεί κατανομή Bern(2/3) με τιμές $\{1, 2\}$. Συνεπώς, αρκεί ο κώδικας $\{0, 1\}$. Στην ουσία έχουμε 2 βιβλία κωδίκων (2 "λεξικά"). Το λεξικό που θα χρησιμοποιήσουμε εξαρτάται από την τιμή της Y, την οποία γνωρίζει τόσο ο συμπιεστής όσο και ο αποσυμπιεστής. Στην περίπτωση όπου $Z = 0$ δε χρειάζεται λεξικό γιατί δεν απαιτείται κωδικοποίηση.

Το μέσο μήκος του κώδικα ισούται με

$$
\begin{aligned}
\mathbb{E}[l] &= \text{Pr}\{Y=0\} \times 0 + \text{Pr}\{Y=1\}\left(\frac{2}{3}\times 1 + \frac{1}{3}\times 1\right) \\
&\quad + \text{Pr}\{Y=2\}\left(\frac{1}{3}\times 1 + \frac{1}{3}\times 2 + \frac{1}{3}\times 2\right) \\
&= \frac{1}{3} + \frac{1}{3}\times\frac{5}{3} = \frac{8}{9} \approx 0.8889 \text{ bits.}
\end{aligned}
$$

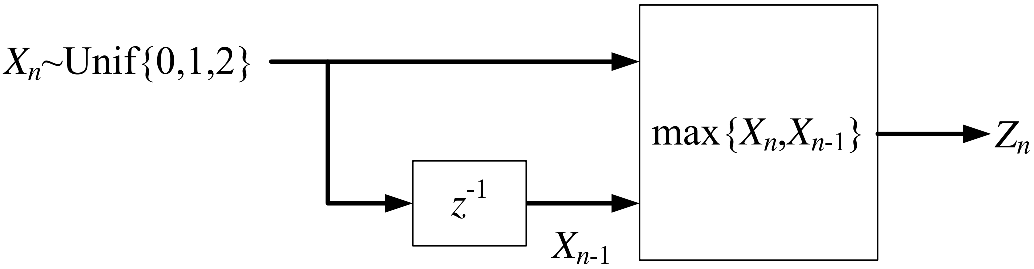
Παρατηρούμε ότι επιτυγχάνεται συμπίεση αρκετά κοντά στην $H(Z|Y) = H(Z|X)$ που υπολογίστηκε στο Ερώτημα (β).

Θεωρήστε, τώρα, το τροποποιημένο σύστημα του Σχήματος 13. Τώρα χρησιμοποιούμε το κύκλωμα max$\{\cdot\}$ σε διαδοχικές χρονικές στιγμές. Η τ.μ. Y έχει αντικατασταθεί από το δείγμα της X την προηγούμενη χρονική στιγμή, δηλαδή $Z_n = \max\{X_n, X_{n-1}\}$. Θεωρούμε, και πάλι, ότι $X_n \sim$Unif$\{0, 1, 2\}$. Επίσης, οι X_n είναι ανεξάρτητες και ομοίως κατανεμημένες (i.i.d.). Θεωρούμε (παρόλο που δεν έχει ιδιαίτερη σημασία) ότι ο χρόνος αρχίζει τη στιγμή $n = 0$ και ότι $X_{-1} = 0$.

(στ) Υπολογίστε την $H(Z_n)$ για $n > 0$ και την $I(X_n; X_{n-1})$ για $n \geq 1$.

Απάντηση:

Η κατανομή της Z_n είναι η ίδια όπως και στο Ερώτημα (α). Επομένως,
$H(Z_n) = \log_2 3$ bits.
Επειδή η $\{X_n\}$ είναι i.i.d., οι X_n και X_{n-1} είναι ανεξάρτητες. Επομένως,
$I(X_n; X_{n-1}) = 0$.



Σχήμα 13: Τροποποιημένο σύστημα.

(*ζ) Υπολογίστε την $H(Z_n|Z_{n-1})$ για $n > 1$. Είναι η $\{Z_n\}$ i.i.d.;

Απάντηση:

Για να βρούμε την $H(Z_n|Z_{n-1})$ παρατηρούμε ότι:

• Αν $Z_{n-1} = 0$ τότε $X_{n-1} = 0$.
Επομένως, $Z_n = \max\{X_n, 0\} = X_n \sim \text{Unif}(\{0,1,2\})$ και $(Z_n|Z_{n-1} = 0) = \log 3$.

• Αν $Z_{n-1} = 1$ τότε $X_{n-1} = 1$ με πιθανότητα 2/3 ή $X_{n-1} = 0$ με πιθανότητα $1/3$ γιατί υπάρχουν 3 ισοπίθανα ενδεχόμενα που παράγουν $Z_{n-1} = 1$: $(X_{n-1}, X_{n-2}) = \{(0,1),(1,0),(1,1)\}$. Με χρήση του θεωρήματος ολικής πιθανότητας,
$p_{Z_n|Z_{n-1}}(Z_n|Z_{n-1} = 1) = \left\{\frac{1}{9}, \frac{5}{9}, \frac{3}{9}\right\}$.
Συνεπώς, $H(Z_n|Z_{n-1} = 1) = H\left(\frac{1}{9}, \frac{5}{9}, \frac{3}{9}\right) \approx 1.3516$ bits.

• Αν $Z_{n-1} = 2$ τότε $X_{n-1} = 0$ με πιθανότητα 1/5 ή $X_{n-1} = 1$ με πιθανότητα $1/5$ ή $X_{n-1} = 2$ με πιθανότητα $3/5$ γιατί υπάρχουν 5 ισοπίθανα ενδεχόμενα που παράγουν $Z_{n-1} = 2$: $(X_{n-1}, X_{n-2}) = \{(0,2),(1,2),(2,2),(2,1),(2,0)\}$. Με χρήση του θεωρήματος ολικής πιθανότητας,
$p_{Z_n|Z_{n-1}}(Z_n|Z_{n-1} = 2) = \left\{\frac{1}{15}, \frac{3}{15}, \frac{11}{15}\right\}$.
Συνεπώς, $H(Z_n|Z_{n-1} = 2) = H\left(\frac{1}{15}, \frac{3}{15}, \frac{11}{15}\right) \approx 1.0530$ bits.

Συνδυάζοντας όλα τα παραπάνω,

$$\begin{aligned} H(Z_n|Z_{n-1}) &= \text{Pr}\{Z_{n-1} = 0\}(Z_n|Z_{n-1} = 0) + \\ &= \text{Pr}\{Z_{n-1} = 1\}(Z_n|Z_{n-1} = 1) + \\ &= \text{Pr}\{Z_{n-1} = 2\}(Z_n|Z_{n-1} = 2) \\ &\approx \frac{1}{9} \times \log 3 + \frac{1}{3} \times 1.3516 + \frac{5}{9} \times 1.0530 = 1.2116 \text{ bits.} \end{aligned}$$

Παρατηρούμε ότι, σε αντίθεση με την περίπτωση όπου οι X και Y προέρχονται από διαφορετικές πηγές (οπότε και $H(Z_n|Z_{n-1}) = H(Z_n) \approx 1.3516$ bits), στην περίπτωση που χρησιμοποιούμε τη X_{n-1} στη θέση της Y η έξοδος δεν είναι, πλέον, i.i.d. Η τελευταία παρατήρηση μπορεί να γίνει και απευθείας από το γεγονός ότι, τόσο η Z_n όσο και η Z_{n-1} εξαρτώνται από τη X_{n-1}.

(*η) Συγκρίνετε την $H(Z_n|Z_{n-1}, Z_{n-2})$ (για $n > 2$) με την $H(Z_n|Z_{n-1})$. Δε χρειάζεται να βρείτε ακριβή τιμή, απλώς να προσδιορίσετε εάν
$H(Z_n|Z_{n-1}, Z_{n-2}) = H(Z_n|Z_{n-1})$ ή όχι. Είναι η $\{Z_n\}$ αλυσίδα Markov 1ης τάξης;

Απάντηση:

Από το προηγούμενο ερώτημα, εάν $Z_{n-1} = 0$, τότε ξέρουμε ότι $X_{n-1} = 0$ με πιθανότητα 1 και γνώση της Z_{n-2} δε μας δίνει καμια επιπλέον πληροφορία που μπορεί να μας χρησιμεύσει για τη Z_n, αφού δεν μπορούμε να ξέρουμε τίποτα από πριν για τη X_n.

Εάν $Z_{n-1} = 1$, αυτό σημαίνει ότι έχει εμφανιστεί ένα από 3 ισοπίθανα ενδεχόμενα (X_{n-1}, X_{n-2}): (0,1), (1,0) ή (1,1). Εάν, τώρα, μάθουμε ότι $Z_{n-2} = 0$, αυτό σημαίνει ότι $X_{n-2} = 0$. Επομένως, μαθαίνουμε με βεβαιότητα ότι $X_{n-1} = 1$. Αυτό αλλάζει την κατανομή της Z_n σε σχέση με την περίπτωση που ξέραμε μόνο τη Z_{n-1}. Άρα, δεδομένου ότι το ενδεχόμενο $(Z_{n-1}, Z_{n-2}) = (1, 0)$ εμφανίζεται με μη μηδενική πιθανότητα, $H(Z_n|Z_{n-1}, Z_{n-2}) \neq H(Z_n|Z_{n-1})$. Γνωρίζουμε, όμως, ότι $H(X|Y) \leq H(X)$. Συνεπώς, $H(Z_n|Z_{n-1}, Z_{n-2}) < H(Z_n|Z_{n-1})$ και η $\{Z_n\}$ δεν είναι αλυσίδα Markov 1ης τάξης!

Παρατηρήσεις:

Γενικά, οι επιδόσεις στην Πρόοδο ήταν καλές. Κάποιοι έχασαν χρόνο για να υπολογίσουν την $p_{Z|X}$ για $X = 0, 1$ και 2 με χρήση του θεωρήματος Bayes. Η προσέγγιση είναι σωστή, απλώς η $p_{Z|X}$ προκύπτει πιο γρήγορα με επισκόπηση του προβλήματος. Το Ερώτημα (ε) προβλημάτισε τους περισσότερους. Κανείς δε μας απαγορεύει να αλλάζουμε τον κώδικα ανάλογα με την τιμή της Y. Ένας κώδικας που εξαρτάται από την τιμή της Y είναι και αυτός ένας κώδικας! Αρκεί κωδικοποιητής και αποκωδικοποιητής να γνωρίζουν ακριβώς τον κώδικα συναρτήσει της Z και της Y.

25. **Αποταμίευση (Τελική Εξέταση Π. Θ. Θ. Π., Ιούνιος 2009)**

Ένας καταθέτης ανοίγει λογαριασμό με αρχικό κεφάλαιο $X_0 = 1000$ και *μηνιαίο* επιτόκιο 1%. Θεωρούμε ότι το επιτόκιο αυτό είναι εγγυημένο για όσο παραμένει ανοικτός ο λογαριασμός, δηλαδή δε μεταβάλλεται. Επίσης, θεωρούμε ότι ο τόκος υπολογίζεται στο τέλος κάθε μήνα. Στο τέλος κάθε μήνα ο καταθέτης έχει την επιλογή να εισπράξει τον τόκο ή να τον αφήσει στο λογαριασμό, οπότε αυτός προστίθεται στο υπάρχον κεφάλαιο. Θεωρούμε, τέλος, ότι δεν επιτρέπεται στον καταθέτη να εισπράξει ποσό διαφορετικό από τον τόκο στο τέλος κάθε μήνα (ούτε μεγαλύτερο ούτε μικρότερο). Δηλαδή ο καταθέτης πρέπει να εισπράξει είτε τον τόκο του μήνα ή τίποτα.

(α) Εάν σε σύνολο N μηνών ο καταθέτης έχει εισπράξει τον τόκο K φορές, δώστε μια έκφραση για το κεφάλαιο, X_N, στο τέλος του $N-$οστού μήνα. Θεωρούμε ότι η X_N ισούται με το κεφάλαιο που απομένει *μετά* από την είσπραξη του τόκου, εφόσον αυτή γίνει. Εάν ο καταθέτης δεν εισπράξει ποτέ τους τόκους, μετά από πόσους μήνες θα έχει διπλασιάσει το αρχικό κεφάλαιο; Δίνεται ότι $1/\log_2(1.01) \approx 69.66$.

Απάντηση:

$$X_N = 1000 \cdot (1.01)^{N-K}.$$

Εάν οι τοκοι δεν εισπραχθούν ποτέ,

$$X_N = 1000 \cdot (1.01)^N = 2000.$$

Λύνοντας ως προς N,

$$N = \lceil \log 2/\log(1.01) \rceil = 70.$$

(β) Θεωρούμε, τώρα, ότι ο καταθέτης ενδέχεται να έχει ανάγκη τους τόκους, με αποτέλεσμα να τους εισπράττει στο τέλος κάθε μήνα με πιθανότητα $1/4$. Η απόφαση αν θα εισπράξει τους τόκους το μήνα i είναι ανεξάρτητη από την απόφασή του το μήνα $j \neq i$.
Με τι ισούται η από κοινού εντροπία $H(X_0, X_1, \ldots, X_N)$;
Με τι ισούται ο ρυθμός εντροπίας, $H(\mathcal{X})$;
Με τι ισούται η από κοινού εντροπία $H(X_0, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N)$, για κάποιο $0 < j < N$;
Δίνεται $\log_2 3 \approx 1.585$.
$\boxed{\text{Απάντηση:}}$
Από τον κανόνα αλυσίδας για την εντροπία,

$$\begin{aligned}
H(X_0, X_1, \ldots, X_N) &= H(X_0) + H(X_1|X_0) + \cdots + H(X_N|X_0, X_1, \ldots, X_{N-1}) \\
&\overset{(i)}{=} H(X_0) + H(X_1|X_0) + H(X_2|X_1) + \cdots + H(X_N|X_{N-1}) \\
&= 0 + \underbrace{H(1/4) + \cdots + H(1/4)}_{\text{όροι}} = NH(1/4) \\
&= N\left(\frac{1}{4}\log 4 + \frac{3}{4}\log\frac{4}{3}\right) = N(2 - \frac{3}{4}\log 3) \approx 0.8113N.
\end{aligned}$$

(i) Οι X_i σχηματίζουν Μαρκοβιανή αλυσίδα.
Ο ρυθμός εντροπίας ισούται με

$$H(\mathcal{X}) = \lim_{N\to\infty} \frac{1}{N} H(X_0, X_1, \ldots, X_N) = H(1/4) \approx 0.8113 \text{ bits}.$$

Εναλλακτικά,

$$H(\mathcal{X}) = \lim_{N\to\infty} H(X_N|X_{N-1}, \ldots, X_0) = H(X_N|X_{N-1}) = H(1/4).$$

Για να υπολογίσουμε την $H(X_0, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N)$ εφαρμόζουμε και πάλι τον κανόνα αλυσίδας.

$$\begin{aligned}
&H(X_0, X_1, \ldots, X_N) = H(X_0, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N) \\
&+ H(X_j|X_0, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N) \Rightarrow \\
&H(X_0, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N) = H(X_0, X_1, \ldots, X_N) \\
&- H(X_j|X_0, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N) \\
&\overset{(ii)}{=} H(X_0, X_1, \ldots, X_N) - H(X_j|X_{j-1}, X_{j+1}).
\end{aligned}$$

(ii) Οι X_i σχηματίζουν Μαρκοβιανή αλυσίδα.
Μπορούμε να υπολογίσουμε την $H(X_j|X_{j-1}, X_{j+1})$ διακρίνοντας περιπτώσεις:

1) $X_{j+1} = X_{j-1}$: Pr$\{X_{j+1} = X_{j-1}\} = (3/4)^2$. Στην περίπτωση αυτή, $X_j = X_{j+1} = X_{j-1}$ και $H(X_j|X_{j-1}, X_{j+1}) = 0$.
2) $X_{j+1} = (1.01)^2 X_{j-1}$: Pr$\{X_{j+1} = (1.01)^2 X_{j-1}\} = (1/4)^2$. Επομένως, $X_j = 1.01 \cdot X_{j-1}$ και $H(X_j|X_{j-1}, X_{j+1}) = 0$.
3) $X_{j+1} = 1.01 \cdot X_{j-1}$: Στην περίπτωση αυτή υπάρχουν δύο ισοπίθανα ενδεχόμενα (το καθένα με πιθανότητα $(1/4) \cdot (3/4)$). $X_{j+1} = X_j = 1.01 \cdot X_{j-1}$ ή $X_{j+1} = 1.01 \cdot X_j = 1.01 \cdot X_{j-1}$. Συνεπώς, $H(X_j|X_{j-1}, X_{j+1}) = 1$.

Συνδυάζοντας τα παραπάνω,

$$H(X_j|X_{j-1}, X_{j+1}) = 2(1/4)(3/4) \cdot 1 = 3/8.$$

Επομένως,

$$H(X_0, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N) = (1/4) - 3/8 \approx 0.8113N - 0.375.$$

Εναλλακτική λύση (δόθηκε από 2 φοιτητές στο διαγώνισμα):
Με χρήση κανόνα αλυσίδας,

$$\begin{aligned}
&H(X_0, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N) \\
&= H(X_0) + H(X_1|X_0) + H(X_2|X_1, X_0) + \ldots \\
&+ H(X_{j-1}|X_0^{j-2}) + H(X_{j+1}|X_0^{j-1}) + H(X_{j+2}|X_{j+1}, X_0^{j-1}) + \ldots + H(X_N|X_{j+1}^N, X_0^{j-1}) \\
&\overset{(iii)}{=} H(X_1|X_0) + H(X_2|X_1) + \ldots + H(X_{j-1}|X_{j-2}) + H(X_{j+1}|X_{j-1}) \\
&+ H(X_{j+2}|X_{j+1}) + \ldots + H(X_N|X_{N-1}) \\
&= (N-2)H(1/4) + H(X_{j+1}|X_{j-1}).
\end{aligned}$$

(iii) Οι X_i σχηματίζουν Μαρκοβιανή αλυσίδα.
Η $H(X_{j+1}|X_{j-1})$ μπορεί να βρεθεί από την $p(x_{j+1}|x_{j-1})$. Εύκολα βλέπουμε ότι $X_{j+1} = \{X_j, (1.01)X_j, (1.01)^2 X_j\}$ με πιθανότητες $\{(1/4)^2, 2(1/4)(3/4), (3/4)^2\}$, αντιστοίχως. Συνεπώς,

$$\begin{aligned}
H(X_{j+1}|X_{j-1}) &= \frac{1}{16}\log 16 + \frac{6}{16}\log\frac{16}{6} + \frac{9}{16}\log\frac{16}{9} \\
&= \frac{1}{4} + \frac{3}{2} - \frac{3}{8}\log 6 + \frac{9}{4} - \frac{9}{16}\log 9 \\
&= \frac{16}{4} - \frac{3}{8} - \frac{3}{8}\log 3 - \frac{9}{8}\log 3 \\
&= \frac{29}{8} - \frac{3}{2}\log 3 \approx 1.2475.
\end{aligned}$$

Επομένως,

$$H(X_0, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N) \approx 0.8113N - 0.375.$$

(γ) Δύο φοιτήτριες προσπαθούν να εκτιμήσουν πόσοι μήνες θα χρειαστούν ώστε ο καταθέτης να καταφέρει να *οκταπλασιάσει* το αρχικό του κεφάλαιο. Η εκτίμηση της πρώτης είναι ότι αυτό θα έχει συμβεί σχεδόν σίγουρα σε 209 μήνες. Η δεύτερη ισχυρίζεται ότι η εκτίμηση αυτή είναι παρακινδυνευμένη και υποθέτει ότι θα πρέπει να περιμένουμε τουλάχιστον 279 μήνες. Ποια από τις δύο εκτιμήσεις είναι ορθότερη; Δικαιολογήστε επαρκώς την απάντησή σας. Όπως και στο προηγούμενο ερώτημα, θεωρούμε ότι, στο τέλος κάθε μήνα, ο καταθέτης εισπράττει τους τόκους με πιθανότητα $1/4$.

Απάντηση:

Η πρώτη φοιτήτρια υποθέτει ότι ο καταθέτης δε θα εισπράξει ποτέ τόκους, ούτως ώστε να ισχύει $1000(1.01)^N > 8000 \Rightarrow N_{\text{min}} = 209$. Ωστόσο, αυτό θα συμβεί με πιθανότητα $(3/4)^{209} \approx 7.72 \cdot 10^{-27}$!

Η δεύτερη φοιτητρια, η οποία πιθανώς γνωρίζει το AEP (ή, ισοδύναμα, το νόμο των μεγάλων αριθμών), σκέφτεται ότι, καθώς το N μεγαλώνει, η πιθανότητα συγκεντρώνεται στις τυπικές ακολουθίες. Επομένως, θεωρεί ότι ο καταθέτης θα εισπράξει τους τόκους *περίπου* $N/4$ φορές και ότι δε θα τους εισπράξει τις υπόλοιπες $3N/4$. Επομένως, επιλύει την εξίσωση $1000(1.01)^{3N/4} > 8000 \Rightarrow N_{\text{min}} = 279$. Παρόλο που ο ακριβής υπολογισμός της πιθανότητας η δεύτερη φοιτήτρια να έχει δίκιο δεν είναι εύκολος, παρατηρούμε, κατ' αρχάς, ότι, δεδομένου ότι περιμέναμε περισσότερο, η πιθανότητα να έχουμε φτάσει τις 8000 είναι μεγαλύτερη από $(3/4)^{209} \approx 7.72 \cdot 10^{-27}$. Ωστόσο, το πόσο κοντά είμαστε στο 1, εξαρτάται από το αν το N είναι αρκετά μεγάλο. Αυτό μπορεί να επαληθευτεί μόνο αριθμητικά και δεν είναι εύκολο.

(*δ) Δώστε μια έκφραση για την τιμή της $H(X_N)$ για $N \to \infty$.

Υπόδειξη: Χρησιμοποιήστε το κεντρικό οριακό θεώρημα. Η έκφραση που θα προκύψει είναι συνάρτηση του N.

Απάντηση:

Έστω Z_i τ.μ. με τιμές 0 και 1 που υποδηλώνει εάν ο καταθέτης εισέπραξε τους τόκους το μήνα i. Επομένως, ο αριθμός των φορών που εισπράχθηκαν οι τόκοι σε διάστημα N μηνών ισούται με $A_N = \sum_{i=1}^{N} Z_i$. Οι Z_i είναι Bernoulli i.i.d. με $p = \Pr\{Z_i = 1\} = 1/4$. Συνεπώς, η ακολουθεί διωνυμική (binomial) κατανομή. Παρατηρούμε ότι $X_N = 1000 \cdot (1.01)^{N-A_N}$. Επομένως, σε κάθε τιμή της A_N αντιστοιχεί μια μοναδική τιμή της X_N η οποία έχει την ίδια μάζα πιθανότητας, με αποτέλεσμα $H(X_N) = H(A_N)$.

Απομένει να βρούμε την εντροπία της διωνυμικής κατανομής. Ωστόσο, αυτό αποτελεί δισεπίλυτο πρόβλημα. Αποδεικνύεται ότι η $H(X_N)$ τείνει ασυμπτωτικά στην τιμή $\frac{1}{2}\log(2\pi e N p(1-p)) + \sum_{k\geq 1} a_k N^{-k}$, όπου a_k σταθερές.

Εμείς θα εστιαστούμε μόνο στον πρώτο όρο. Από το κεντρικό οριακό θεώρημα γνωρίζουμε ότι ο μέσος όρος αθροίσματος i.i.d. τ.μ. τείνει στη Γκαουσιανή κατανομή με μέσο όρο μ και διασπορά σ^2/N, όπου μ και σ^2 ο μέσος όρος και η διασπορά, αντιστοίχως, της . Δεδομένου ότι η εντροπία συνεχούς τ.μ. δεν εξαρτάται από το μέσο όρο της, εξετάζουμε τη διασπορά. Εύκολα προκύπτει ότι η διασπορά τ.μ. Bernoulli ισούται με $\sigma^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p(1-p)$. Συνεπώς, $[((1/N)\sum Z_i)^2 - \mu^2] = p(1-p) \Rightarrow \mathbb{E}[(\sum Z_i)^2 - N^2\mu^2] = N^2 p(1-p)$. Επομένως, ή A_N τείνει σε Γκαουσιανή κατανομή με μέση τιμή μ και διασπορά $N^2 p(1-p)/N = Np(1-p)$.

Αντικαθιστώντας στην έκφραση για τη (διαφορική) εντροπία Γκαουσιανής τ.μ.,

$$H(X_N) = H(A^N) \to \frac{1}{2}\log(2\pi e N p(1-p)) \to \frac{1}{2}\log(N) \text{ για } N \to \infty.$$

Εναλλακτική λύση (από Α. Μεσολογγίτη):

Μπορούμε να γράψουμε $X_N = 1000 \prod_{i=1}^{N}(1+0.01Z_i)$, όπου Z_i τ.μ. Bernoulli που υποδηλώνει εάν ο τόκος εισπράχτηκε ($Z_i = 0$) ή όχι. Λογαριθμίζοντας,

$$\log(X_N) = \log(1000) + \sum_{i=1}^{N} \log(1+0.01Z_i).$$

Εάν θέσουμε $S_N \triangleq X_N - 1000$ και $Y_i \triangleq \log(1+0.01Z_i)$, η S_N είναι άθροισμα i.i.d. τ.μ. Y_i με μέση τιμή

$$\begin{aligned}
[\log(1+0.01Z_i)] &= p \cdot 0 + (1-p) \cdot \log 1.01 = 0.0108, \text{ και} \\
\sigma^2(Y_i) &= [\log(1+0.01Z_i)^2] - (0.0108)^2 \\
&= p \cdot 0 + (1-p) \log(1.01)^2 - (0.0108)^2 = 3.864 \cdot 10^{-5}.
\end{aligned}$$

Και πάλι, επειδή υπάρχει αντιστοιχία 1-προς-1 μεταξύ $S_N = \log(X_N)$ και X_N,

$$H(X_N) = H(S_N) \to \frac{1}{2}\log(2\pi e N \times 3.864 \cdot 10^{-5}) \to \frac{1}{2}\log(N) \text{ για } N \to \infty.$$

Η διαφορά εδώ είναι ότι η S_N δεν ακολουθεί διωνυμική κατανομή γιατί οι Y_i δεν είναι Bernoulli.

26. **Τυπικές ακολουθίες (Επαναληπτική Εξέταση Π. Θ. Θ. Π., Σεπτέμβριος 2010)**

Θεωρούμε πηγή *χωρίς μνήμη* που παράγει ακολουθία δυαδικών τ.μ. $X_i \sim$Bern(p). Δηλαδή, η $X_1, X_2, \ldots$ είναι ακολουθία ανεξάρτητων και ομοίως κατανεμημένων (i.i.d.) τ.μ.

(α) Εάν $\left|A_\epsilon^{(n)}\right|$ είναι το πλήθος των ασθενώς $\epsilon-$τυπικών ακολουθιών μήκους n και $|X_1^n|$ είναι όλες οι δυαδικές ακολουθίες μήκους n, τι μπορούμε να πούμε για το λόγο $\left|A_\epsilon^{(n)}\right| / |X_1^n|$ καθώς $n \to \infty$ και $\epsilon \to 0$;

Υπόδειξη: Πρέπει να δείτε αν ο λόγος συγκλίνει στην ίδια τιμή για όλες τις τιμές της παραμέτρου p.

Απάντηση:

Γνωρίζουμε ότι $|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$. Επομένως,

$$\frac{\left|A_\epsilon^{(n)}\right|}{|X_1^n|} \leq \frac{2^{n(H(X)+\epsilon)}}{2^n} = 2^{n(H(X)-1+\epsilon)}.$$

Δεδομένου ότι, για δυαδικές τ.μ., $H(X) < 1$ όταν $p \neq \frac{1}{2}$, ο λόγος συκλίνει στο 0. Δηλαδή, το πλήθος των τυπικών ακολουθιών είναι αμελητέο σε σύγκριση με όλες τις δυαδικές ακολουθίες. Εξαίρεση αποτελεί η περίπτωση $p = \frac{1}{2}$, όπου όλες οι ακολουθίες είναι (ασθενώς) τυπικές.

(β) Θεωρούμε, τώρα, ότι οι τιμές της X_i είναι 0 (με πιθανότητα p) ή 1 (με πιθανότητα $1-p$). Ορίζουμε το βάρος Hamming (Hamming weight), $W(X_1^n)$, της ακολουθίας X_1^n ως τον αριθμό των '1' της ακολουθίας.
Δείξτε ότι $\frac{1}{n}W(X_1^n) \to 1-p$ για $n \to \infty$.

Υπόδειξη: Χρησιμοποιήστε το νόμο των μεγάλων αριθμών.

Απάντηση:

Από το νόμο των μεγάλων αριθμών, στο όριο η ακολουθία θα περιέχει pn '0' και $(1-p)n$ '1'. Επομένως, $W(X_1^n) \to n(1-p) \Rightarrow \frac{1}{n}W(X_1^n) \to 1-p$.

(γ) Έστω, τώρα, ότι το μήκος, n, της ακολουθίας είναι πεπερασμένο και ότι $\epsilon > 0$. Δείξτε ότι μπορούμε να αποφασίσουμε αν η ακολουθία X_1^n μήκους n είναι *ασθενώς* $\epsilon-$τυπική αν γνωρίζουμε την τιμή του βάρους Hamming, $W(X_1^n)$, της ακολουθίας.

Απάντηση:

Για να είναι η ακολουθία X_1^n ασθενώς $\epsilon-$τυπική πρέπει να ισχύει

$$H(X) - \epsilon \leq -\frac{1}{n} \log p(x_1, x_2, \ldots, x_n) \leq H(X) + \epsilon$$

Εάν η ακολουθία περιέχει r '1' και $n-r$ '0', μπορούμε να γράψουμε

$$\begin{aligned}
&H(X) - \epsilon \leq -\frac{1}{n} \log \prod_{i=1}^{n} p(x_i) \leq H(X) + \epsilon \Rightarrow \\
&H(X) - \epsilon \leq -\frac{1}{n} \log \left(p^{n-r}(1-p)^r\right) \leq H(X) + \epsilon \Rightarrow \\
&H(X) - \epsilon \leq -\frac{1}{n} \log \left(p^n \left(\frac{1-p}{p}\right)^r\right) \leq H(X) + \epsilon \Rightarrow \\
&H(X) - \epsilon \leq -\log p - \frac{r}{n} \log \left(\frac{1-p}{p}\right) \leq H(X) + \epsilon \Rightarrow \\
&H(X) - \epsilon + \log p \leq \frac{r}{n} \log \left(\frac{p}{1-p}\right) \leq H(X) + \epsilon + \log p \Rightarrow \\
&n\frac{H(X) - \epsilon + \log p}{\log(p/(1-p))} \leq r \leq n\frac{H(X) + \epsilon + \log p}{\log(p/(1-p))}.
\end{aligned}$$

Παρατηρούμε ότι $r = W(X_1^n)$. Επομένως, με βάση το διάστημα στο οποίο βρίσκεται το βάρος Hamming της ακολουθίας μπορούμε να προσδιορίσουμε αν η ακολουθία είναι ασθενώς $\epsilon-$τυπική.