ΑΣΚΗΣΗ 1η

Υπολογισμός εντροπίας πηγής πληροφορίας χωρίς μνήμη

1.1 Σκοπός της άσκησης

Σκοπός της άσκησης είναι να υπολογιστεί η *εντροπία* (entropy) μιας πηγής πληροφορίας χωρίς μνήμη δύο δυαδικών συμβόλων και να πραγματοποιηθεί η αναπαράστασή της σε συνάρτηση με την πιθανότητα εμφάνισης των συμβόλων.

1.2 Θεωρητικό μέρος

Ας θεωρήσουμε μία πηγή πληροφορίας στην οποία οι πιθανότητες εμφάνισης των συμβόλων της είναι ανεξάρτητες του χρόνου δηλαδή θεωρούμε μία *στατική πηγή* πληροφορίας (πηγή χωρίς μνήμη). Οι πηγές πληροφορίας χωρίς μνήμη, χαρακτηρίζονται από την εντροπία τους $\mathrm{H}_{\text{πηγή χωρίς μνήμη}}$ ή μέση κατά σύμβολο πληροφορία που δίνεται από τη παρακάτω σχέση, [1-3]:

$$\mathrm{H}_{\text{πηγή χωρίς μνήμη}} = \sum_{i=1}^{n} \mathrm{p}_i \cdot \log_2\left(\frac{1}{\mathrm{p}_i}\right) \tag{1.1}$$

όπου στη σχέση (1.1), n είναι το πλήθος των συμβόλων της πηγής και p_i η πιθανότητα εμφάνισης του αντιστοίχου συμβόλου i.

Ας θεωρήσουμε μία πηγή χωρίς μνήμη δύο μόνο συμβόλων. Αν η πιθανότητα εμφάνισης του ενός συμβόλου είναι p, του άλλου συμβόλου θα είναι (1-p). Συνεπώς σύμφωνα με τη σχέση (1.1), η εντροπία της συγκεκριμένης πηγής πληροφορίας θα δίνεται από τη σχέση:

$$\mathrm{H}_{\text{πηγή δύο συμβόλων}} = p \cdot \log_2\left(\frac{1}{\mathrm{p}}\right) + (1-p)\cdot \log_2\left(\frac{1}{1-\mathrm{p}}\right) \tag{1.2}$$

Η εντροπία της προηγούμενης πηγής γίνεται *μέγιστη* όταν οι πιθανότητες εμφάνισης των δύο συμβόλων είναι ίσες δηλαδή όταν τα δύο σύμβολα της πηγής είναι ισοπίθανα. Η προηγούμενη παρατήρηση ισχύει γενικά για οποιαδήποτε πηγή πληροφορίας δηλαδή η σχέση (1.1) παρουσιάζει μέγιστο (μέγιστο της εντροπίας) όταν όλα τα n σύμβολα της πηγής είναι ισοπίθανα, [2].

1.3 Εργαστηριακό μέρος

Να γράψετε στο MATLAB πρόγραμμα υπολογισμού και αναπαράστασης της εντροπίας πηγής πληροφορίας χωρίς μνήμη δύο συμβόλων. Να παρατηρήσετε αν η εντροπία έχει μέγιστη τιμή και ποια είναι αυτή η τιμή. Να εξηγήσετε το αποτέλεσμα. Η αναπαράσταση να πραγματοποιηθεί σε συνάρτηση με την πιθανότητα εμφάνισης p του ενός από τα δύο σύμβολα της πηγής πληροφορίας.

Πρόγραμμα

```
function H=Entropy(p)
echo off
p=[0:0.1:1];
for i=1:11
H(i)=-p(i)*log2(p(i))-(1-p(i))*log2(1-p(i));
```

```
End

pause % Press a key to see a plot of entropy versus probability of binary alphabet
clf
plot(p, H), grid on;
xlabel('Probability p of the binary symbol');
ylabel('Binary Entropy H ');
```

Εξήγηση Προγράμματος

Στην αρχή του προγράμματος απαιτούμε να μην εμφανίζονται πάλι στην έξοδο οι εντολές που εκτελούνται χρησιμοποιώντας την εντολή echo off. Είναι γνωστό, από τη θεωρία, ότι η πιθανότητα εμφάνισης *p* ενός συμβόλου μιας πηγής πληροφορίας, λαμβάνει τιμές μεταξύ του 0 και του 1. Συνεπώς στο πρόγραμμα υπολογισμού της εντροπίας μιας πηγής πληροφορίας χωρίς μνήμη δύο συμβόλων, θα πρέπει να ορίσουμε μία μεταβλητή p η οποία θα λαμβάνει τις αντίστοιχες τιμές. Αυτό πραγματοποιείται μέσω της εντολής:

p=[0:0.1:1]; (1.3)

Είναι προφανές ότι δίνουμε διακριτές τιμές στη μεταβλητή p διότι στη συνέχεια θα προχωρήσουμε σε αντίστοιχη γραφική αναπαράσταση. Το βήμα αύξησης της μεταβλητής p θεωρήσαμε ότι είναι ίσο με 0.1. Θα μπορούσαμε να θεωρήσουμε και μεγαλύτερη τιμή για το βήμα αύξησης της μεταβλητής p, αλλά θα πρέπει να προχωρήσουμε στη συνέχεια σε γραφική παράσταση η οποία θα πρέπει να έχει ικανοποιητικές τιμές για τις δύο μεταβλητές (δηλαδή να είναι ικανοποιητικό το πλήθος των τιμών των δύο αξόνων x και y).

Με δεδομένο ότι η μεταβλητή p λαμβάνει διακριτές τιμές, ορίσαμε μία συνάρτηση H η οποία θα αναπαρασταίνει την εντροπία της πηγής και η οποία θα λαμβάνει αντίστοιχα διακριτές τιμές. Αυτό πραγματοποιείται μέσω της εντολής:

H(i)=-p(i)*log2(p(i))-(1-p(i))*log2(1-p(i)); (1.4)

Με δεδομένο το βήμα αύξησης της πιθανότητας εμφάνισης του συμβόλου p, να είναι ίσο με 0.1 και τις τιμές της από 0 έως και 1, η εντροπία της πηγής H θα λαμβάνει διακριτές τιμές οι οποίες θα πρέπει να υπολογίζονται κάθε φορά για την αντίστοιχη τιμή της πιθανότητας p. H διαδικασία αυτή απαιτεί τη δημιουργία ενός *βρόχου επανάληψης* (loop). Ο βρόχος επανάληψης θα εκτελείται, με δεδομένο το βήμα αύξησης της πιθανότητας p να είναι ίσο με 0.1, έντεκα φορές. Σε κάθε εκτέλεση θα υπολογίζεται μία τιμή για την εντροπία H αντίστοιχη της τιμής της πιθανότητας εμφάνισης του συμβόλου p. H προηγούμενη διαδικασία πραγματοποιείται συνολικά με τον παρακάτω βρόχο επανάληψης:

```
for i=1:11
H(i)=-p(i)*log2(p(i))-(1-p(i))*log2(1-p(i));
End
```
(1.5)

H εντολή:

pause % Press a key to see a plot of entropy versus probability of binary alphabet(1.6)

προχωρεί στη φάση της γραφικής παράστασης της εντροπίας H σε συνάρτηση με την πιθανότητα p μετά από το πάτημα ενός κουμπιού.

Η γραφική παράσταση της εντροπίας H σε συνάρτηση με την πιθανότητα εμφάνισης του ενός συμβόλου p, πραγματοποιείται με τη βοήθεια της εντολής *plot(x,y):*

plot(p, H), grid on; (1.6)

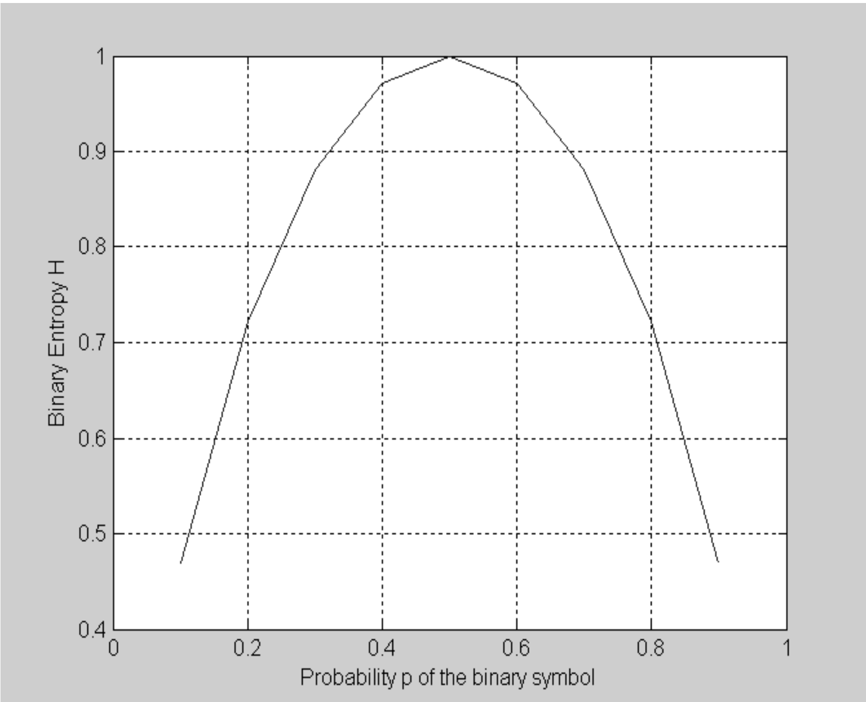
Επίσης, έχουμε ζητήσει από το MATLAB να εμφανίσει "πλέγμα" (*grid on*) στη γραφική παράσταση που θα παράγει.

Τέλος, στην γραφική παράσταση της εξόδου δίνουμε ονομασίες (label) ("λεζάντες") στους δύο άξονες x και y με τις εντολές:

xlabel('Probability p of the binary symbol');
ylabel('Binary Entropy H '); (1.7)

Έξοδος προγράμματος

Η έξοδος του προγράμματος που αναλύθηκε προηγουμένως, στο MATLAB, [4], παρουσιάζεται στην επόμενη εικόνα 1.1.



Εικόνα 1.1 *Εντροπία πηγής χωρίς μνήμη* (Binary entropy, H) δύο συμβόλων σε συνάρτηση με την *πιθανότητα εμφάνισης του ενός δυαδικού συμβόλου* p (Probablity of the binary symbol, p)*(Έξοδος από εκτέλεση του προγράμματος στο MATLAB).*

.

1.4 Πρόσθετες εργασίες

1. Από την έξοδο του προγράμματος να βρείτε για ποια τιμή της πιθανότητας εμφάνισης του συμβόλου p η εντροπία της πηγής γίνεται μέγιστη και να βρείτε την αντίστοιχη μέγιστη τιμή της.
2. Να γράψετε πρόγραμμα υπολογισμού της εντροπίας πηγής χωρίς μνήμη στην οποία εκπέμπονται τρία σύμβολα με αντίστοιχες πιθανότητες εμφάνισης 0.2, 0.3, 0.5.