Πόση Πληροφορία;

Βασισμένο στο P. Verlinde, *Information Theory,* από Encycl. Of Information Systems

Μια και μιλάμε για Πληροφοριακά Συστήματα που επεξεργάζονται και διακινούν πληροφορία, ας δούμε πώς μπορούμε να τη μετρήσουμε. Ας συμφωνήσουμε ότι:

- Η ποσότητα της πληροφορίας ενός γεγονότος (πχ, ενός μηνύματος) εξαρτάται από το βαθμό της αβεβαιότητας του περιεχομένου του.
- Ένα μήνυμα περιέχει πολλή πληροφορία, όταν το περιεχόμενό του είναι απίθανο. Πχ, «το ποτάμι στέρεψε το καλοκαίρι» περιέχει περισσότερη πληροφορία από το «χιόνισε το Σεπτέμβρη».
- Αν δυο μηνύματα είναι άσχετα μεταξύ τους, η συνολική πληροφορία είναι το άθροισμα των πληροφοριών του κάθε μηνύματος ξεχωριστά.

Λαμβάνοντας αυτά υπόψη, ο Claude Shannon, το 1948 πρότεινε η **πληροφορία** $I(a)$ ενός γεγονότος a να υπολογίζεται ως

$$I(a) = -\log(P(a))$$

όπου $P(a)$ είναι η πιθανότητα να συμβεί το γεγονός a. Επειδή $P(a) \in [0,1] \Rightarrow \log\big(P(a)\big) \geq 0$, άρα η μέτρηση της πληροφορίας μπορεί να πάρει θετικές τιμές από το 0 και πάνω.

Για δυο ανεξάρτητα γεγονότα α και β, η πιθανότητα να συμβούν και τα δύο είναι $P(\alpha \cap \beta) = P(\alpha) \cdot P(\beta)$, οπότε η πληροφορία και των δυο μαζί είναι

$$\begin{aligned} I(\alpha \cap \beta) &= -\log P(\alpha \cap \beta) = -\log\big(P(\alpha) \cdot P(\beta)\big) \\ &= -(\log(P(\alpha) + \log(P(\beta)) \\ &= (-\log(P(\alpha)) + (-\log(P(\beta))) \\ &= I(\alpha) + I(\beta) \end{aligned}$$

δηλαδή ίση με το άθροισμα της πληροφορίας των επιμέρους γεγονότων, όπως είπαμε παραπάνω.

Όταν υπολογίζουμε την πληροφορία χρησιμοποιώντας ως βάση του λογαρίθμου το 2, η **μονάδα μέτρησης** ονομάζεται bit.[1]

Παράδειγμα: Αν ρίξεις ένα τίμιο κέρμα στον αέρα, το μήνυμα «το κέρμα έφερε γράμματα» έχει πιθανότητα 0,5 και άρα πληροφορία $-\log_2 1/2 = 1\,bit$. Το μήνυμα «το κέρμα έφερε κορώνα ή γράμματα» έχει πιθανότητα 1 και άρα πληροφορία $-\log_2 1 = 0\,bit$.

Αν ρίξεις n κέρματα, η πιθανότητα να φέρουν όλα γράμματα είναι $\frac{1}{2} \cdot \frac{1}{2} \cdots \frac{1}{2} = \frac{1}{2^n}$ και άρα η πληροφορία αυτού του γεγονότος είναι $n\,bits$. Γενικότερα λοιπόν, αν έχουμε ένα μήνυμα που μεταφέρει ένα γεγονός από 2^n ισοπίθανες παραλλαγές του, τότε το μήνυμα αυτό περιέχει $n\,bits$ πληροφορίας. Αυτό το αποτέλεσμα ‘συμφωνεί’ με τη δυαδική αναπαράσταση που χρησιμοποιούμε στους υπολογιστές, πχ για $n = 3$, οι οκτώ παραλλαγές είναι οι 000, 001, 010, 011, 100, 101, 110, 111 και στέλνοντας μια από αυτές απαιτούνται 3 δυαδικά ψηφία. (Κατά σύμπτωση, η μονάδα μέτρησης της πληροφορίας, το bit, έχει το ίδιο όνομα με τα ψηφία της δυαδικής αναπαράστασης.) Συνεπώς, δεν υπάρχει πιο οικονομικός τρόπος να μεταδώσουμε το μήνυμα ότι όλα τα κέρματα έφεραν γράμματα από μια ακολουθία μήκους n —εφόσον οι παραλλαγές είναι ισοπίθανες, δηλαδή τα κέρματα είναι τίμια.

Αντί να υπολογίζουμε την πληροφορία ενός συγκεκριμένου γεγονότος από τη γκάμα αυτών που μπορεί να συμβούν, είναι πιο χρήσιμο να υπολογίσουμε το μέσο όρο της πληροφορίας όλων των πιθανών γεγονότων. Αυτή η ποσότητα ονομάζεται **εντροπία.** Άρα η εντροπία $H(X)$ ενός γεγονότος X που έχει πολλά τυχαία αποτελέσματα x, ορίζεται ως

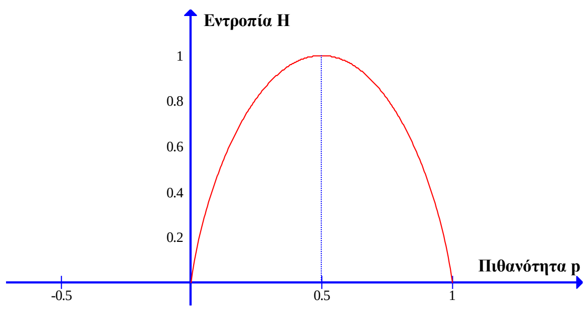
$$H(X) = \sum_{\forall x} P(x) \cdot I(x)$$

όπου $P(x)$ είναι η πιθανότητα να συμβεί το x και $I(x)$ είναι η πληροφορία του x. Άρα

$$H(X) = -\sum_{\forall x} P(x) \cdot log(P(x))$$

Η εντροπία είναι ένα μέτρο της αβεβαιότητας του γεγονότος. Παίρνει τιμές στο διάστημα $[0,1]$.

Παράδειγμα: Στη γενική περίπτωση όπου το νόμισμα δεν είναι τίμιο, η πιθανότητα να φέρει γράμματα είναι p και η πιθανότητα να φέρει κορώνα είναι $1 - p$. Τότε μεταβολή της εντροπίας του φαίνεται στο γράφημα



Παρατηρήστε ότι η εντροπία λαμβάνει τη μέγιστη τιμή της, $H = 1$, όταν $p = \frac{1}{2}$, δηλαδή όταν το ζάρι είναι τίμιο και πράγματι τότε είναι πιο αβέβαιο το αποτέλεσμα.

[1] Εγκυκλοπαιδικά, όταν η βάση του αλγορίθμου είναι το 10, η μονάδα μέτρησης λέγεται Hartley και όταν είναι e, λέγεται Shannon ή nat.