Complex Systems Methods — 1. Probability, Entropy and Information

Eckehard Olbrich

MPI MIS Leipzig

Potsdam WS 2007/08

Overview

1. Probability
   - Discrete random variable
   - Continuous random variable
2. Entropy
   - Uniqueness of entropy
   - Properties of the entropy
   - Conditional entropy
   - Mutual information
   - Properties of the Mutual Information
   - Differential entropy

Interpretations of probability

- Objective probabilities: Probabilities describe a property of the "objective world" and are measured by relative frequencies — frequentist
- Subjective probabilities: Probabilites describe the degree of uncertainty about the occurence of an event — Bayesian
- Kolmogorov: Probability is a non-negative measure normalized to unity on a σ-algebra of elementary events

Random variable

Probability space $(\Omega, \mathcal{A}, P)$

Set of possible events Ω: Set of outcomes of an random experiment — in the case of a coin toss $\Omega = (\mathit{heads}, \mathit{tails})$. Elements denoted by $\omega \in \Omega$.

σ-algebra of subsets $\mathcal{A}$: Set of subsets of Ω.

Probability measure P: Each set of events $A \subseteq \mathcal{A}$ has a probability $0 \leq P(A) \leq 1$. $P(\Omega) = 1$.

Random variable X

Measureable function $X : (\Omega, \mathcal{A}) \to S$ to a measurable space S (frequently taken to be the real numbers with the standard measure). The probability measure $PX^{-1} : S \to \mathbb{R}$ associated to the random variable is defined by $PX^{-1}(s) = P(X^{-1}(s))$. A random variable has either an associated probability distribution (discrete random variable) or probability density function (continuous random variable).

Discrete random variable

A random variable X is said to be *discrete* if the set $\{X(\omega) : \omega \in \Omega\}$ (i.e. the range of X) is finite or countable.

Alphabet: Set $\mathcal{X}$ of values of the random variable X.

Probability: $p(x) = P(X = x), x \in \mathcal{X}$.

Normalization:

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

Expectation value of X:

$$E_P[X] = \sum_{x \in \mathcal{X}} x p(x)$$

Continuous random variable

A random variable X is said to be continuous if it has a cumulative distribution function which is absolutely continuous.

Probability density $p(x)$

$$P(a \leq X \leq b) = \int_a^b p(x)dx \; .$$

Cumulative distribution

$$P_{\leq}(x) = P(X \leq x) = \int_{-\infty}^{x} p(y)dy$$

Normalization

$$\int_{x_{min}}^{x_{max}} p(x)dx = 1 \; .$$

Expectation value

$$E[X] = \int_{-\infty}^{\infty} x p(x) dx$$

Median $x_{1/2}$

$$P_{\leq}(x_{1/2}) = \frac{1}{2}$$

Change of variable $y = f(x)$ (f invertible)

$$p(x)dx = q(y)dy \quad \Rightarrow \quad q(y) = \left. \frac{p(x)}{df/dx} \right|_{x=f^{-1}(y)}$$

Entropy

- Shannon 1948: How much choice is involved in the selection of an event with n possibilities and probabilities $p_1, \ldots, p_n$?
- If we have a random variable X with a probability distribution $p(x)$ the uncertainty about the outcome x of a measurement of X is given by the *entropy*
- $$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) \ .$$
- Entropy can be considered as a measure of variety or disorder ("objective") or as a measure of uncertainty ("subjective")
- Information reduces uncertainty, i.e. it can be quantified by differences between uncertainties, that is: entropies.
- The entropy can be considered as the expextation value of $\log 1/p(x)$:

$$H(X) = E_P[\log \frac{1}{p(x)}] \ .$$

Uniqueness of entropy

Are there other functions, which are suitable as a measure of uncertainty?

Theorem: The following three conditions determine the function $H(p_1, \ldots, p_n)$ up to a multiplicative constant, whose value serves only to determine the size of the unit of information.

1. $H(p, 1-p)$ is a continuous function of $p \in [0, 1]$.
2. $H_n(p_1, \ldots, p_n)$ is a symmetric function of all of its arguments.
3. If $p_n = q_1 + q_2 > 0$ then

$$H(p_1, p_2, p_3, \ldots, q1, q2) = H(p_1, p_2, p_3, \ldots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right) .$$

The last property called "additivity" is dropped for some entropies such as the *Renyi entropies*.

Properties of the entropy

1. $H(X) \geq 0$, because $0 \leq p(x) \leq 1$ implies that $\log 1/p(x) \geq 0$.
2. $H_b(X) = (\log_b a) H_a(X)$, i.e. entropy in nats $H_e(X) = (\ln 2) H_2(X)$ with $H_2(X)$ the entropy in bits, because $\log_b p = \log_b a \log_a p$.
3. Binary random variable, $X = 0$ with p and $X = 1$ with $1 - p$.

$$H(p) := H(X) = -p \log p - (1-p) \log(1-p) \ .$$

$H(p) = 0$ for $p = 0$ and $p = 1$ and $H(p)$ maximal for $p = 1/2$.

Conditional entropy

- Knowing Y might reduce the uncertainty about X if both are not statistically independent.
- The uncertainty of X having already observed $Y = y$ can be expressed as

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

- This can be averaged also over Y giving

$$H(X|Y) \quad = \quad H(X, Y) - H(Y) \; .$$

  $H(X|Y)$ is called *conditional entropy*.
- Chain rule:

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

Mutual information

The information, .i.e. the reduction of unvertainty, about X provided by knowing already Y can be quantified using the mean difference between the uncertainty of X knowing Y and without knowing Y:

$$\begin{aligned} MI(X:B) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X,Y) \\ &= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \end{aligned}$$

This difference is called the mutual information between X and Y. It can be considered also as a measure of *correlation* or statistical dependence between X and Y.

Properties of the Mutual Information

Definition: The *relative entropy* or *Kullback-Leibler divergence* between two distributions $p(x)$ and $q(x)$ is defined as

$$\begin{aligned} D(p||q) &= \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \\ &= E_p \left[\frac{p(x)}{q(x)} \right] \end{aligned}$$

Proposition: The relative entropy is a non-negative quantity $D(p||q) \geq 0$. It becomes zero if and only if $p = q$.

Mutual information: The mutual information can be expressed as the Kullback-Leibler divergence between the joint distribution $p(x, y)$ and product of its marginals $p(x)p(y)$.

$$MI(X:Y) = D(p(x,y)||p(x)p(y)) \, .$$

Corollary: The mutual information is non-negative:

$$MI(X:Y) \geq 0 \, .$$

Differential entropy

- Entropy of a continous random variable X (also *continuius entropy*)

$$h(X) = -\int_{-\infty}^{\infty} p(x) \log p(x)$$

- The differential entropy might be negative.
- The differential entropy depends on the scale of measurement $h(aX) = h(X) + \log |a|$.
- If the density $p(x)$ of the random variable X is Riemann integrable, then

$$H(X^{\epsilon}) + \log \epsilon \rightarrow h(X) \text{ , as } \epsilon \rightarrow 0 \text{ .}$$