3F1 - Signals and Systems

Michaelmas 2005

Information Theory
Handout 4

Andrea Lecchini Visintini

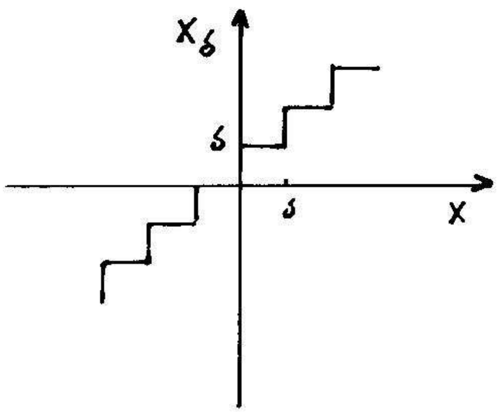
30 November 2005

The differential Entropy

The **differential Entropy** (or Entropy of a continuous variable) is a generalization of the concept of Entropy for random variables with a continuous distribution. The differential Entropy of a continuous variable X is derived by calculating the Entropy of a discrete variable X_δ, which is obtained by quantizing X, and then by taking the limit as the size of the quantization step tends to zero. In contrast with the Entropy of discrete variables, the differential Entropy turns out to be a **relative** quantity which can be positive or negative.

Let X be a continuous random variable, with probability density function $f(x)$, and let X_δ be a discrete random variable obtained by rounding off X as

$$X_\delta = n\delta \qquad n\,(\delta - 1) < X \le n\,\delta\,.$$



The distribution of probability of X_δ is given by

$$P\{X_\delta = n\delta\} = P\{n\,(\delta-1) < X \le n\delta\} = \int_{n\,(\delta-1)}^{n\,\delta} f(x)\,dx$$
$$= \delta\bar{f}(n\delta)$$

where $\bar{f}(n\delta)$ is a number between the maximum and the minimum of $f(x)$ in the interval $[n\,(\delta-1),\ \ n\delta]$. The Entropy of X_δ is

$$H(X_\delta) = -\sum_{n=-\infty}^{\infty} \delta\bar{f}(n\delta)\log_2[\delta\bar{f}(n\delta)]$$
$$= -\log_2(\delta)\sum_{n=-\infty}^{\infty}\delta\bar{f}(n\delta) \;-\; \sum_{n=-\infty}^{\infty}\delta\bar{f}(n\delta)\log_2[\bar{f}(n\delta)]$$
$$= -\log_2(\delta) \;-\; \sum_{n=-\infty}^{\infty}\delta\bar{f}(n\delta)\log_2[\bar{f}(n\delta)]\,.$$

(here we assume the convention that $f(x)\log_2[f(x)] = 0$ if $f(x) = 0$). Notice that $H(X_\delta)$ tends to infinity as δ tends to 0. This is consistent with the fact that it would take an infinite number of bits to specify the value of a continuous value with arbitrary precision.

The **differential Entropy** of the continuous random variable X is defined on the basis of the above calculations and is given by

$$H(X) = \lim_{\delta\to 0}[H(X_\delta) + \log_2(\delta)] = -\int_{-\infty}^{\infty} f(x)\log_2[f(x)]\,dx\,.$$

The idea behind the definition of differential Entropy, which justifies the elimination of $\log_2(\delta)$, is that usually we are interested in the **difference** (hence the name) in the values of Entropy between random variables, rather that in the Entropy of a single random variable, and therefore the term $\log_2(\delta)$ is irrelevant.

The term "Entropy of X" is commonly used to denote the "differential Entropy of X" when X is a continuous variable. Sometimes the convention is also to choose the natural logarithm in the formula of the Entropy of continuous distributions. Here we will continue to use the base 2.

Examples

1) Let X have uniform density in $[0,a]$ then

$$H(X) = \log_2 a\,.$$

Notice that the uniform distribution defined on $[0,1]$ has Entropy = 0. Hence it can be considered as attaining a "reference level" for the differential Entropies of all the other densities.

2) Let X have Laplace density $f(x) = \frac{1}{2b}e^{-\frac{|x-\mu|}{b}}$ then

$$\begin{aligned}
H(X) &= -\frac{1}{\ln(2)}\int_{-\infty}^{\infty} f(x)\ln\left(\frac{1}{2b}e^{-\frac{|x-\mu|}{b}}\right)dx \\
&= -\frac{1}{\ln(2)}\int_{-\infty}^{\infty} f(x)\left[-\frac{|x-\mu|}{b} - \ln(2b)\right]dx \\
&= \frac{1}{\ln(2)}\int_{-\infty}^{\infty} f(x)\frac{|x-\mu|}{b}dx + \frac{\ln(2b)}{\ln(2)} \\
&= \frac{1}{\ln(2)}\frac{1}{b}\int_{0}^{\infty} \frac{x}{b}\, e^{-\frac{x}{b}}dx + \frac{\ln(2b)}{\ln(2)} \\
&= \frac{1}{\ln(2)} + \frac{\ln(2b)}{\ln(2)} = \log_2(2eb)
\end{aligned}$$

3) Let $X_{\mathcal{N}}$ have Normal density $n(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ then

$$\begin{aligned}
H(X_{\mathcal{N}}) &= -\frac{1}{\ln(2)}\int_{-\infty}^{\infty} n(x)\ln\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)dx \\
&= -\frac{1}{\ln(2)}\int_{-\infty}^{\infty} n(x)\left[-\frac{(x-\mu)^2}{2\sigma^2} - \ln\left(\sigma\sqrt{2\pi}\right)\right]dx \\
&= \frac{1}{\ln(2)}\frac{1}{2\sigma^2}\int_{-\infty}^{\infty} n(x)(x-\mu)^2dx + \frac{\ln\left(\sigma\sqrt{2\pi}\right)}{\ln(2)} \\
&= \frac{1}{\ln(2)}\frac{1}{2\sigma^2}E[(X_{\mathcal{N}}-\mu)^2] + \frac{\ln\left(\sigma\sqrt{2\pi}\right)}{\ln(2)} \\
&= \frac{1}{\ln(2)}\frac{1}{2} + \frac{\ln\left(\sigma\sqrt{2\pi}\right)}{\ln(2)} \\
&= \log_2(\sigma\sqrt{2\pi e})
\end{aligned}$$

The joint and conditional Entropy and the mutual Information are defined in a similar way as the discrete case. Let X and Y have joint distribution $f(x,y)$ then:

the **joint Entropy** is given by

$$H(X,Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log_2[f(x,y)]dx\,dy\,,$$

the **conditional Entropy** is

$$H(Y|X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log_2[f(y|x)]dx\,dy\,,$$

the **mutual Information** is

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

$$\begin{aligned}
&H(X,Y) \le H(X) + H(Y) \\
&H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \\
&H(X|Y) \le H(X)
\end{aligned}$$

The MaxEnt principle

The Entropy can be used as a criterion to create probability models which embed available information with fairness. The typical problem is that of creating a density function which is consistent with given values of some parameters which in turn are not sufficient to determine a unique density. These values may have been found through experiments and one wants to construct a probabilistic model which is consistent with the experimental results. The **MaxEnt principle** states that the least biased model which embeds the available information is that which maximizes the Entropy while remaining consistent with this information.

In the following Theorem we show that the Normal density is the MaxEnt density among all the densities with given values of mean and variance.

Theorem
The Normal density $\mathcal{N}(\mu, \sigma^2)$ is the distribution which maximizes the Entropy among all the possible densities which give mean μ and variance σ^2. □

Proof
Let $X_{\mathcal{N}}$ have density $\mathcal{N}(\mu, \sigma^2)$, which is denoted $n(x)$, and let $X_{\mathcal{G}}$ have a generic density $g(x)$ with $E[X_{\mathcal{G}}] = \mu$ and $E[(X_{\mathcal{G}} - \mu)^2] = \sigma^2$. To start with, we have that

$$H(X_{\mathcal{N}}) = -\frac{1}{\ln(2)} \int_{-\infty}^{\infty} g(x) \ln\left[n(x)\right] dx\,.$$

In fact on the right hand side we have that

$$\begin{aligned}
\int_{-\infty}^{\infty} g(x) \ln\left[n(x)\right] dx &= \int_{-\infty}^{\infty} g(x) \left[-\frac{(x-\mu)^2}{2\sigma^2} - \ln\left(\sigma\sqrt{2\pi}\right)\right] dx \\
&= \frac{1}{2\sigma^2}\int_{-\infty}^{\infty} g(x)(x-\mu)^2 dx + \ln\left(\sigma\sqrt{2\pi}\right) \\
&= \frac{1}{2\sigma^2}\cdot\sigma^2 + \ln\left(\sigma\sqrt{2\pi}\right)\,.
\end{aligned}$$

Hence we obtain

$$\begin{aligned}
H(X_{\mathcal{G}}) - H(X_{\mathcal{N}}) &= -\frac{1}{\ln(2)}\left[\int_{-\infty}^{\infty} g(x)\ln[g(x)]dx + \int_{-\infty}^{\infty} g(x)\ln[n(x)]dx\right] \\
&= \frac{1}{\ln(2)}\int_{-\infty}^{\infty} g(x) \ln\left[\frac{n(x)}{g(x)}\right] dx \\
&\leq \frac{1}{\ln(2)}\int_{-\infty}^{\infty} g(x) \left[\frac{n(x)}{g(x)} - 1\right] dx \\
&= \frac{1}{\ln(2)}\int_{-\infty}^{\infty} n(x) - g(x) dx \\
&= 0\,.
\end{aligned}$$

□