Let me make up a brief and informal summary about what I know on measuring/calculating entropy.

DISCRETE DISTRIBUTIONS

The original definition of Shannon entropy or informational entropy is on discrete distributions. $H=-\sum_{i=1}^{N} p_i \ln(p_i)$ Where p_i is the probability of *i* class. This is a well defined quantity with well known and amazing properties.

From the point of view of calculation/measure $H=-\langle \ln(p_i)\rangle$ that is, the minus log probability mean. This is a well defined quantity for each discrete probability distribution function (pdf), e.g. A binomial with parameters size=N, probability=p has an entropy $H_{B(N,p)}=\ln\left(\sqrt{2\pi e\, N\, p(1-p)}\right)$.

From a practical point of view what you need to "measure" entropy is a sample from a population. Each discrete pdf defines only one "natural" histogram, with each class represented with a bin. With the trick $0\cdot\ln(0)=0$ entropy could be computed as $\hat{H}=-\sum_{i=1}^{N} p_i \ln(p_i)=\ln(m_S)-\sum_{i=1}^{N} q_i \ln\left(\frac{q_i}{m_S}\right)$ where q_i is the number of elements of the sample in the *i* bin and $m_S=\sum_{i=1}^{N} q_i ;\quad p_i=\frac{q_i}{m_S}$ N is the number of bins of this histogram. This is the Maximum Likelihood estimator. A lot of work has been done to improve the statistical properties of this estimator and a bunch of alternative estimators is available. In the following let stand $\hat{H}$ for any of this convenient estimators.

CONTINUOUS DISTRIBUTIONS

There is a number of reasons to wish the entropy concept working on continuous pdf . The natural way to do this is see the sum of the original definition as an integral $H=-\int_{-\infty}^{\infty} f(x)\ln(f(x))\,dx$ where *f(x)* is the pdf of the random continuous variable *x*. This is known as differential entropy. This is a well defined quantity provided we are integrating a pdf. But... this integral could be negative. What means negative entropy? Problems.

From the point of view of calculation/measure $H=-\langle \ln(p_i)\rangle$ that is, the minus log probability mean. This is a well defined quantity for each probability distribution function (pdf), e.g. A Normal with parameters standard deviation=σ, mean doesn't matter, has an entropy $H_{N(\mu,\sigma)}=\ln\left(\sqrt{2\pi e\sigma^2}\right)$ .But... the standard deviation has the same units than the mean, and the argument of the log function must be dimensionless. Problems. This lead us to relative entropies and the Kullback-Lieber framework $D_{KL}=-\int_{-\infty}^{\infty} f(x)\ln\left(\frac{f(x)}{g(x)}\right)dx$ where *f* and *g* are pdf. This framework works fine in a wide range of situations, but this is a function of two pdf.

From a practical point of view the first thing you need to "measure" entropy is a sample from a population. Applying the standard trick of approximate integrals by sums we recover the ML estimator $\hat{H}=-\sum_{i=1}^{N} p_i \ln(p_i)=\ln(m_S)-\sum_{i=1}^{N} q_i \ln\left(\frac{q_i}{m_S}\right)$ so we only need build up an histogram and apply the estimator. But... a continuous pdf has infinite "natural" histograms, which choose? Each histogram gives us a different estimate, what does it mean? Problems.

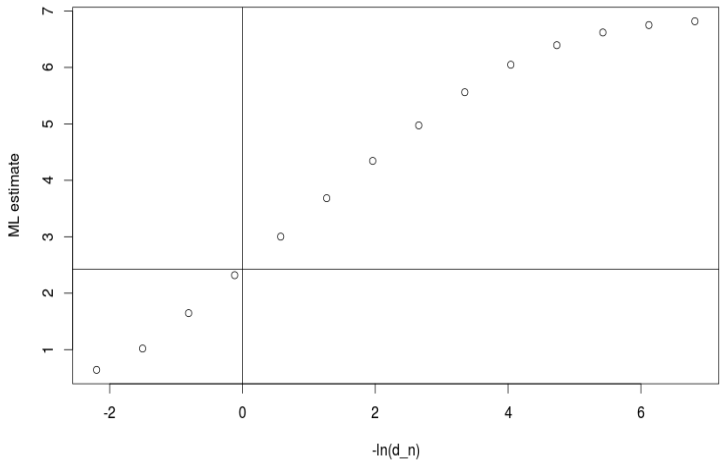
So the three points of view lead us to problems and to an alternative to Shannon entropy. It seems the Shannon entropy doesn't exists or isn't defined for continuous pdf.

Let me try another approach.

First I'll try to solve "Each histogram gives us a different estimate, what does it mean?" Well it means that the entropy estimate depends on "the histogram". The best suited histograms for this approach are the equally distributed. Thus each histogram is defined for the number of bins, *N* in the following way: define $d_N = \frac{Max(sample) - min(sample)}{N}$ the size of the bin. So the *k* element of the sample belongs to the *i-th* bin iff $(i-1)d_N < e_k < i\,d_N$ That's all that I need to apply one of the entropy estimators. How does the entropy estimation vs. $-\ln(d_N)$ look like and what does it means? I think it means that the continuous entropy is not an absolute measure, it is somehow referred to the units scale. **So the entropy measures the uncertainty at the units.** This idea enables us to make some tricks to solve problems:

1. I can transform any dimensional argument to its adimensional version dividing it by 1 unit (a constant, not a function as in the Kullback approach). It makes sense if entropy is referred to the units scale.
2. A negative entropy means a uniform uncertainty volume less than 1 unit of volume. So this is not catastrophic.

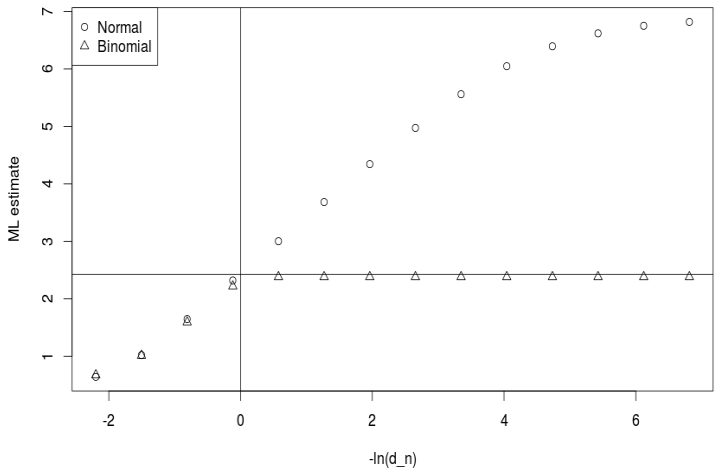
To see how looks like entropy estimation vs. $-\ln(d_N)$ we can draw a sample of size 1000 from a Normal with mean=15,sd=2.73861266846244 and *H*=2.42639 and the following series of *N* the number of bins of each histogram: (2 , 4, 8,16, 32, 64,128,256,512,1024,2048,4096, 8192, 16384), so we have a vector $d_N = \frac{Max(sample) - min(sample)}{(2,4,8,16,...)}$ I'll use the ML estimator as $\hat{H}$ and we get one entropy estimation $\hat{H}_i$ for each histogram. Plotting $\hat{H}_i$ *v.s.* $-\ln(d_N)$ we get



It resembles the Box-Counting algorithm. The horizontal line is at $\ln(\sqrt{2\pi e}\,\sigma)$ in this case 2.42639, the entropy value expected.

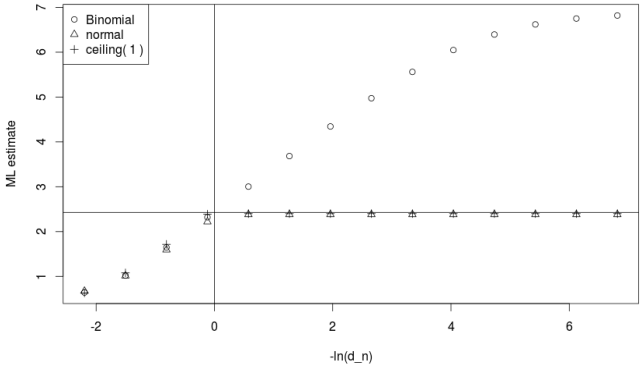
This is the basis of the ebc algorithm.

But, what about the discrete pdf? Let me put the points of the estimation in the same graph. What pdf use? Well, must be a numerical discrete random variable. A standard procedure is see a Binomial of size *N* and probability *p* as a Normal with mean *N·p* and variance *N·p·(1-p)*. That is Binomial tends to Normal. In this case $B(N=30, p=0.5) \leftrightarrow N(\mu=15, \sigma=2.73861266846244)$ It is worth to say that these pdf are isoentropic, that is $\ln(\sqrt{2\pi e\, N\, p(1-p)}) = \ln(\sqrt{2\pi e}\,\sigma) = 2.42639$



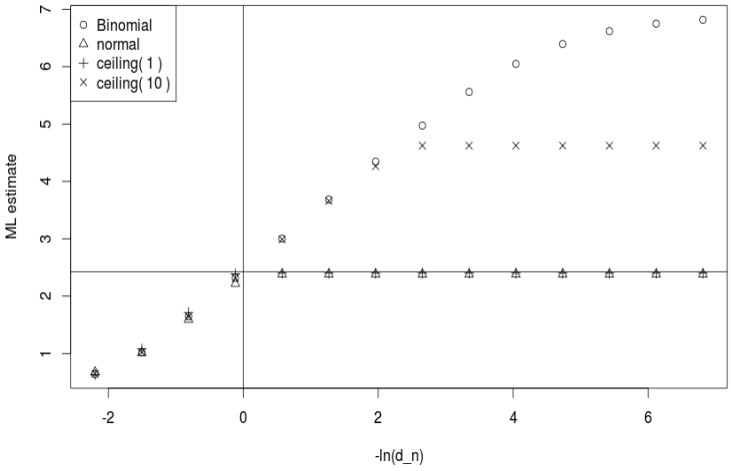
Well, as expected the discrete pdf has a plateau at the correct value of entropy. Of course this plateau means thas there is no more information at this scales. It seems that the pdfs share a portion of the graph.

Let me discretize the continuous sample by *new sample=ceiling(old sample)* and plot it in the same graph



It is clear that ceiling(1) sample behaves the same that the binomial sample.

What if we discretize in a lazy fashion, taking rationals instead of integers? Say *new sample=ceiling((old sample)·10)/10* and let me plot it in the same graph



A new plateau appears. The plateau begins approx at $d_N = \exp(-2.3) \approx 0.1$ That is, when there is not more information. But the entropy is shifted too to a value of 4.62468, but it is a no-sense; this sample must by isoentropic to the others by construction. On the other hand, this sample shares the graph with the continuous sample until the plateau begins.

So the entropy that I want, that I compute is the entropy when $\ln(d_N) = 0$ that is: the entropy at the scale of units.

Summarizing if we take the point of view that entropy is referred to the scale of our units, the units in which I measure or describe the system under study, then we can make a coherent estimation of the entropy from samples of continuous variables. This point of view also allow us to have only one Shannon entropy concept that works for discrete and continuous pdf.