Differential entropy examples: information theory lecture 7

COMSM0075 Information Processing and Brain

`comsm0075.github.io`

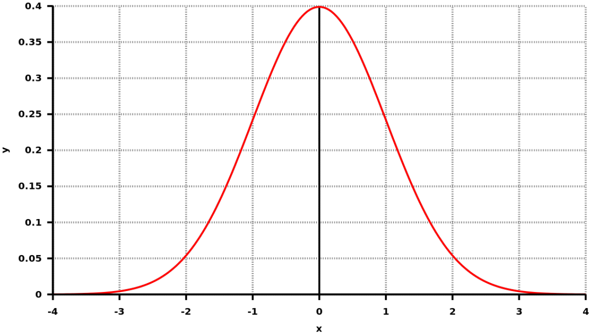
October 2020

Differential entropy

Differential entropy is the name given to Shannon's entropy for continuous probability distributions: distributions where the sample space is $\mathcal{X} \subseteq \mathbf{R}^d$.

Differential entropy

Differential entropy is the name given to Shannon's entropy for continuous probability distributions.

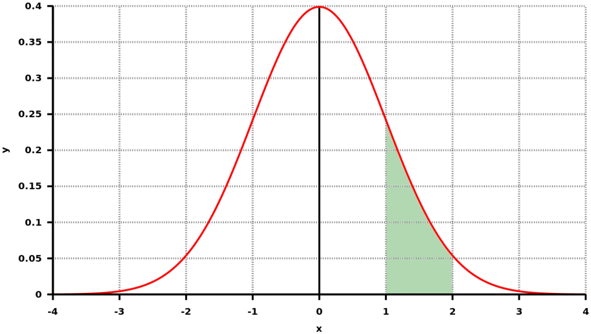


$p(x)$

Image from wikipedia.

Differential entropy

Differential entropy is the name given to Shannon's entropy for continuous probability distributions.



$$P(x_0 \leq x < x_1) = \int_{x_0}^{x_1} dx p(x)$$

is the probability $x_0 \leq x < x_1$.

Image from wikipedia.

Differential entropy

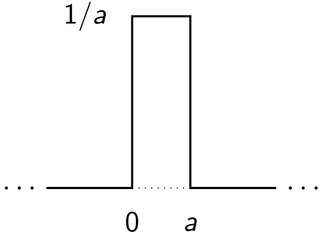
Differential entropy is the name given to Shannon's entropy for continuous probability distributions

$$h(X) = -\int dx p(x) \log_2 p(x)$$

Example - uniform

Consider a uniform distribution

$$p(x) = \begin{cases} 1/a & x \in [0, a] \\ 0 & \text{otherwise} \end{cases}$$



Example - uniform

Consider a uniform distribution

$$p(x) = \begin{cases} 1/a & x \in [0, a] \\ 0 & \text{otherwise} \end{cases}$$

so

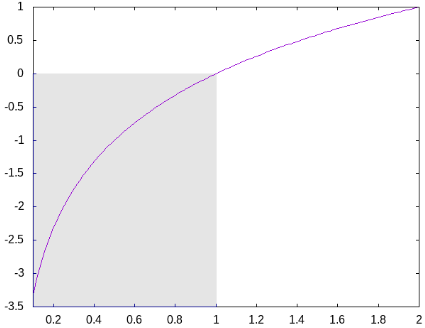
$$h(X) = -\int_{-\infty}^{\infty} dx\, p(x) \log_2 p(x) = -\frac{1}{a}\int_0^a dx\, \log_2 \frac{1}{a}$$

and so

$$h(X) = \log_2 a$$

Who ordered that!

$$h(X) = \log_2 a$$



The Gaußian distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$$

Substitute and integrate by parts to get

$$h(X) = \frac{1}{2} \log_2 2\pi e\sigma^2$$

where the e is just the exponential $\exp(1)$.

More of the negative values

As with the uniform distribution, this formula can give a positive or negative number depending on the size of σ*. Interestingly it can be proved that for fixed variance the Gaußian has the highest entropy.*

Densities are not probabilities

discrete case $p(x)$ is the probability of $X = x$

continuous case

$$\int_{x_0}^{x_1} dx p(x)$$

is the probability $x_0 \leq x < x_1$.

Densities are not probabilities

The usual sums and probabilities go to integrals and densities doesn't work because there is a p in the log:

$$h(X) = -\int dx p(x) \log_2 p(x)$$

Real numbers are like an infinity of numbers

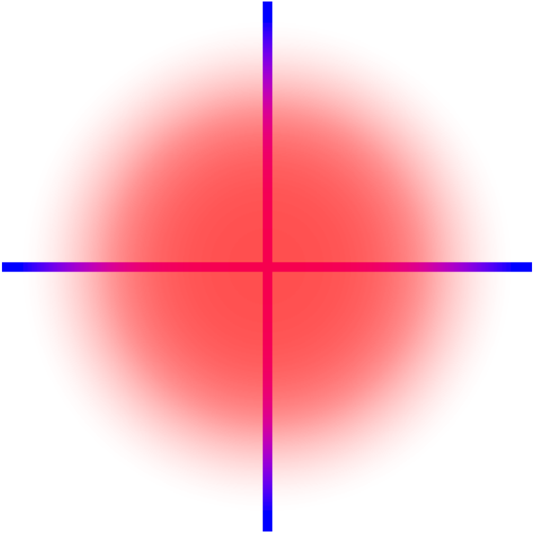
1.618033988749894848204586834365638117720309179805762862135
44862270526046281890244970720720418939113748475408807538689
17521266338622235369317931800607667263544333890865959395829
05638322661319928290267880675208766892501711696207032221043
21626954862629631361443814975870122034080588795445474924618
56953648644492410443207713449470495658467885098743394422125
44877066478091588460749988712400765217057517978834166256249
40758906970400028121042762177111777805315317141011704666599
14669798731761356006708748071013179523689427521948435305678
30022878569978297783478458782289110976250030269615617002504
64338243776486102838312683303724292675263116533924731671112
11588186385133162038400522216579128667529465490681131715993
43235973494985090409476213222981017261070596116456299098162
90555208524790352406020172799747175342777592778625619432082
7505131218156285512224809394...

We can't really see all the numbers

1.618033988749894848204586834365638117720309179805762862135
44862270526046281890244970720720418939113748475408807538689
17521266338622235369317931800607667263544333890865959395829
05638322661319928290267880675208766892501711696207032221043
21626954862629631361443814975870122034080588795445474924618
56953648644492410443207713449470495658467885098743394422125
44877066478091588460749988712400765217057517978834166256249
40758906970400028121042762177111777805315317141011704666599
14669798731761356006708748071013179523689427521948435305678
30022878569978297783478458782289110976250030269615617002504
64338243776486102838312683303724292675263116533924731671112
11588186385133162038400522216579128667529465490681131715993
43235973494985090409476213222981017261070596116456299098162
90555208524790352406020172799747175342777592778625619432082
75051312181562855122248093947...

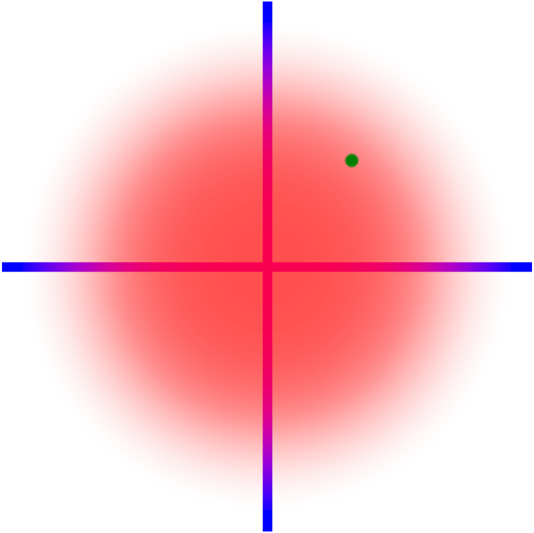
Sensitivity

We need to model the sensitivity of the receiver as well the behaviour of the source.



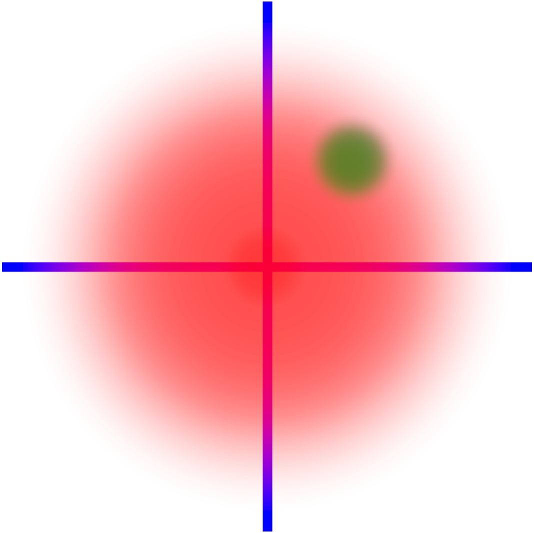
Sensitivity

We need to model the sensitivity of the receiver as well the behaviour of the source.



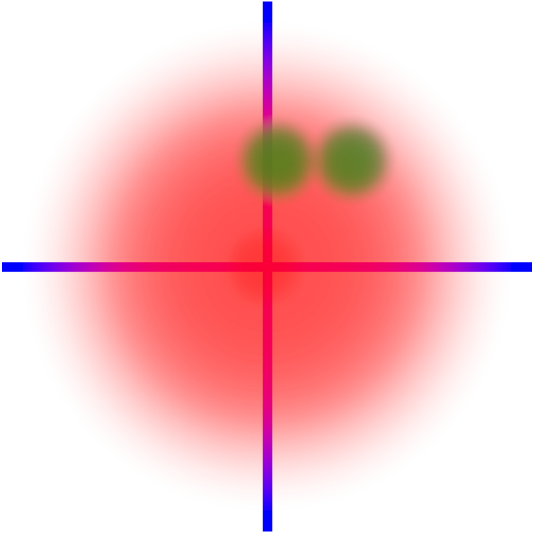
Sensitivity

We need to model the sensitivity of the receiver as well the behaviour of the source.



Sensitivity

We need to model the sensitivity of the receiver as well the behaviour of the source.



Sensitivity

Maybe $I(X, Y)$ is what we really want.