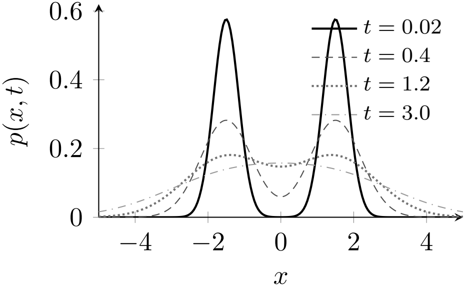
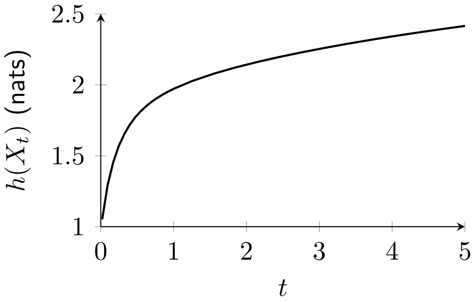
Entropy in Discrete and Continuous Spaces and a Relation to the Heat Equation

Huy

Euler Circle IRPW, July 9, 2026



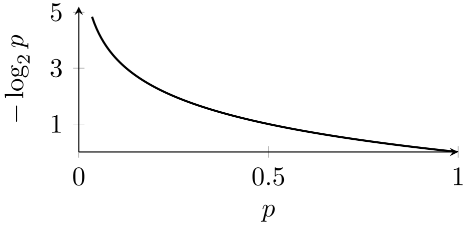


Surprise

k digits name 2^k outcomes $\qquad$ N equally likely $\Rightarrow$ $\log_2 N$ digits

a bit is one binary digit, and every digit you add doubles what you can name

surprise of an outcome of probability $p = -\log_2 p$



probabilities multiply, surprises add, hence the log · rare = expensive

Entropy

$$H(X) \;=\; -\sum_{i} p_i \log_2 p_i$$

entropy = expected surprise, each outcome weighted by how often it happens

Source coding theorem (informal): lossless codes can average $H(X)$ bits per symbol, never fewer. (Shannon 1948)

$\underbrace{\texttt{TTTTTTTT}}_{\text{almost always T}} \;\longrightarrow\; H \approx 0$ $\qquad$ $\underbrace{\texttt{XTTTXTXT}}_{\text{fair coin}} \longrightarrow 1$ bit/symbol

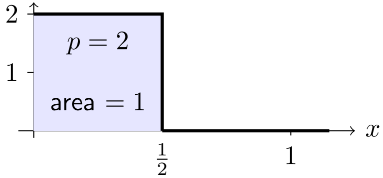
English: $\;4.7 \;\longrightarrow\; 4.14 \;\longrightarrow \approx 1$ bit per letter

measured by people guessing text one character at a time (Shannon 1951)

Differential entropy

$$h(X) \;=\; -\int p \ln p$$

the same formula, sum to integral; areas under p are probabilities. Natural log from here on.

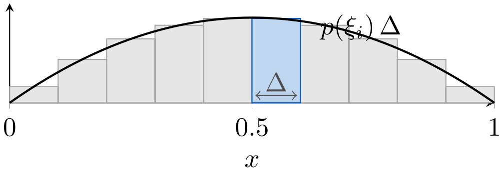


$$h \;=\; -\ln 2 \;\approx\; -0.69 \;<\; 0$$

$-\ln p$ is the constant $-\ln 2$, and the average of a constant is that constant

Expected surprise can't go negative. So what is h measuring?

Measuring at resolution Δ



the bins make a discrete distribution, bin i carries the shaded area $p(\xi_i)\Delta$

$$H_\Delta \approx h + \ln\frac{1}{\Delta}$$ the entropy of the bins (split: paper §5.1)

$$\log_2\frac{1}{\Delta} = \text{binary digits to name the bin} \longrightarrow \infty$$ halve Δ, add one digit

$H_\Delta \to \infty$: a continuous variable has no finite entropy. h is the finite part.

What is it measuring?

What h is

$$h = -\int p \ln p = -\int p \ln \frac{p}{m}, \qquad m \equiv 1$$

p compared, point by point, against the flat density: height one everywhere. Not an amount, a comparison.

$h > 0$: more spread than unit scale $\qquad$ $h < 0$: more concentrated

the uniform on $[0, 1]$ sits exactly at $h = 0$. And $-\ln 2$ means: one bit cheaper than the unit interval, at every resolution.

The flat reference

$\int m = \infty :$ m is not a probability distribution

so the comparison is not forced ≥ 0, and the divergent $\ln \frac{1}{\Delta}$ was the reference's infinity, never p's

h was never a continuous H. It is p measured against the flat reference.

Relative entropy

$$\sum_i p(\xi_i)\Delta \ln \frac{p(\xi_i)\,\Delta}{q(\xi_i')\,\Delta} \;\longrightarrow\; \int p \ln \frac{p}{q} \;=\; D(p\,\|\,q)$$

two real densities at one resolution: the Δs cancel before any limit. No $\ln \frac{1}{\Delta}$ to remove, no correction.

$$D(p\,\|\,q) \;\geq\; 0, \qquad = 0 \text{ only when } q = p$$

nonnegative for a genuine density q, and zero exactly when there is nothing to compare

The fundamental object

discrete was the same: $H(p) \;=\; \ln n \;-\; D(p \,\|\, u)$

reference minus comparison, the same split. $\ln n$ is finite, so on an alphabet it hides inside one number.

Relative entropy is the fundamental object. H **is the case** $q = u$**;** h **is the case** $q = m \equiv 1$**.**

The heat equation

$$\partial_t p \;=\; \tfrac{1}{2}\,\partial_x^2 p$$

peaks sink, valleys fill: ink spreading in water, heat evening out along a bar

$$\text{point mass} \;\longrightarrow\; \mathcal{N}(0,t), \qquad h\big(\mathcal{N}(0,t)\big) = \tfrac{1}{2}\ln(2\pi e t)$$

the heat kernel: width $\sqrt{t}$, spreading forever, entropy growing

$$X_t \;=\; X_0 + \sqrt{t}\,Z$$

running the flow for time $t =$ adding Gaussian noise of variance t. Always growing? At what rate?

de Bruijn's identity

$$J(p) \;=\; \int \frac{(p')^2}{p}$$

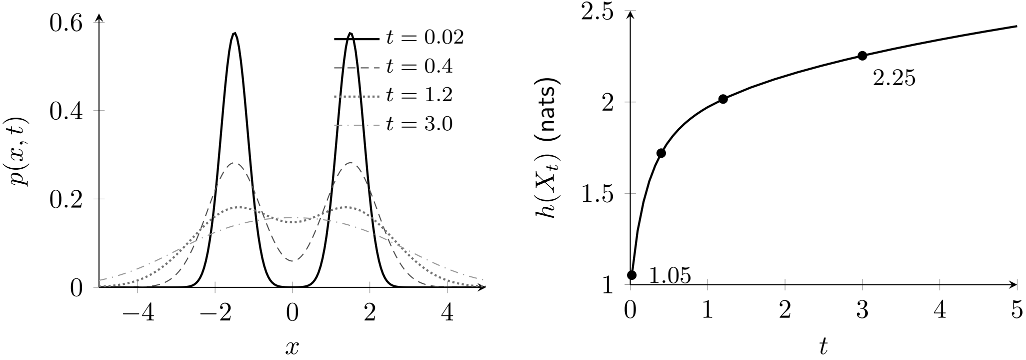
sharp features $\longrightarrow$ large J

$$\boxed{\frac{d}{dt}\,h(p_t) \;=\; \tfrac{1}{2}\,J(p_t)}$$

Stam (1959), crediting de Bruijn

$J \geq 0$, so entropy never falls. Fast while sharp features remain, slower as the density smooths.

The picture



the slope on the right, at every instant, is $\frac{1}{2}J$ of the density on the left