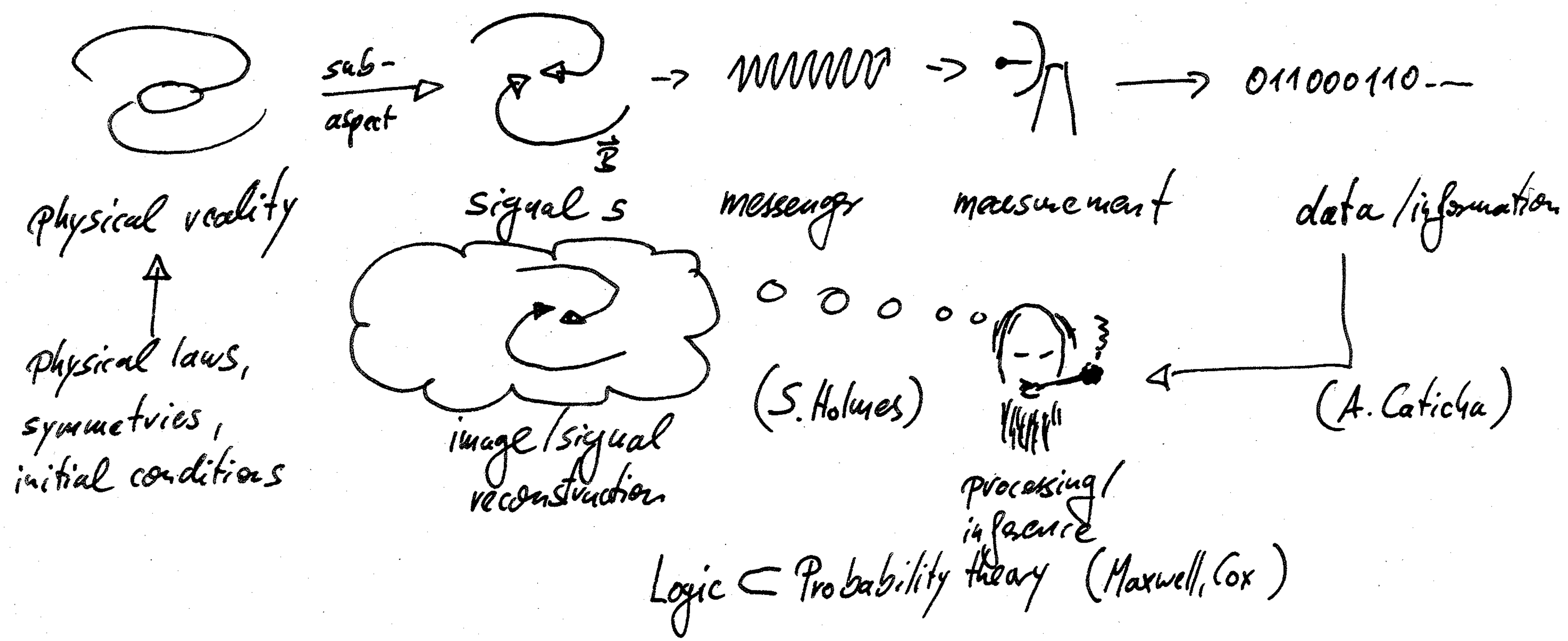
# Data, Signal, Information and all that – Torsten Enßlin, Ringberg, July 2011 -1-



<u>Quotes</u>

Sherlock Holmes / Sir Arthur Conan Doyle : „Most people, if you describe a train of events to them, will tell you what the result will be. They can put those events together in their minds, and argue from them that something will come to pass..." ← <u>commentary on simulators</u>

„... There are a few people, however, who, if you told them a result, would be able to evolve from their own inner consciousness what the steps were that had led to ~~that~~ the result. This power is what I mean when I talk of reasoning backwards, or analytically."
← <u>commenting on scientists</u>

Ariel Caticha 2008 : "Information is what changes rational believes"

Maxwell 1850 : „The actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on.
Therefore the true logic ~~science~~ for this world is the calculus of probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind"

<u>more quotes</u>

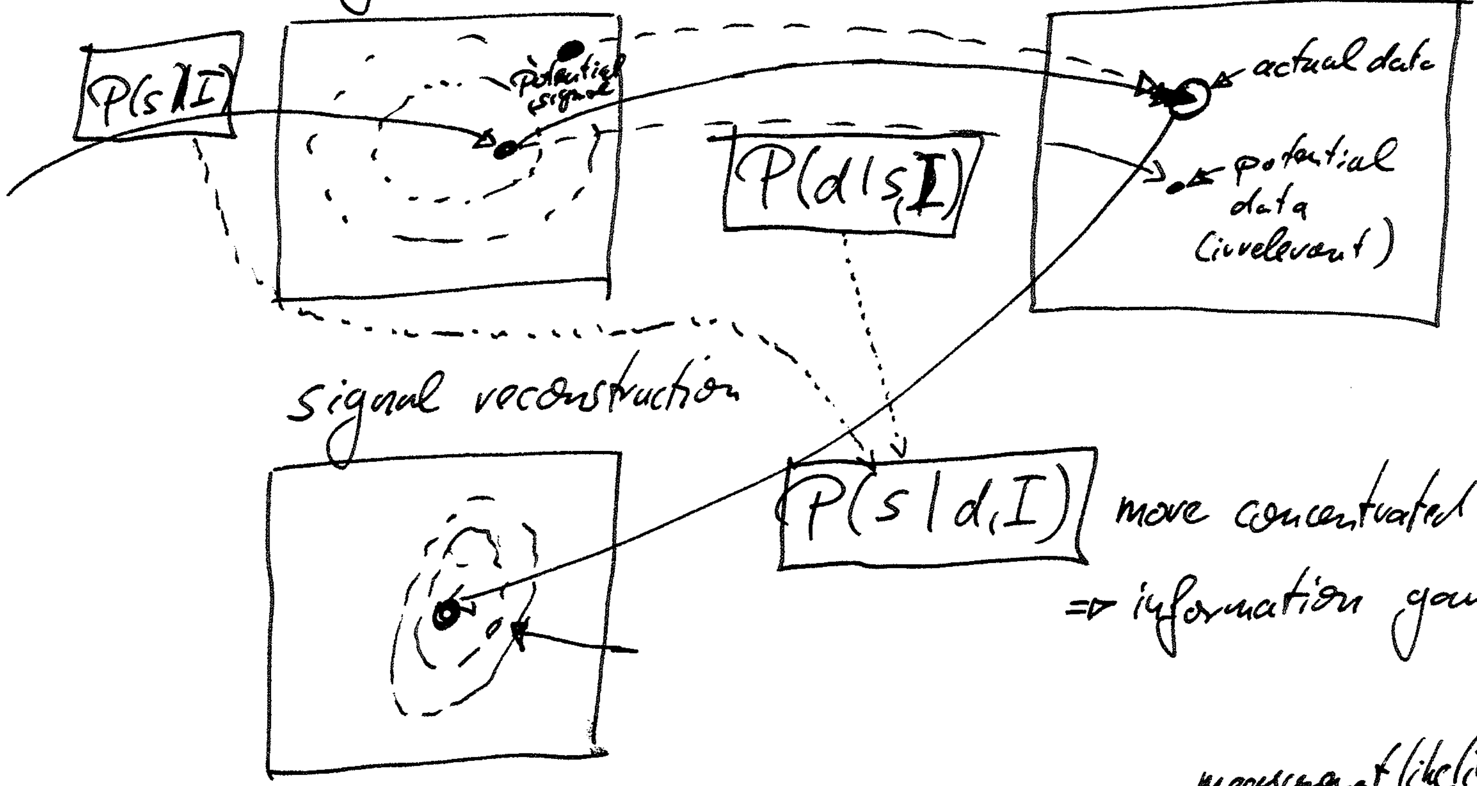
R.T. Cox's theorem 1946: probability theory is the logic of uncertainty

Jaynes 1957: <u>Maximum entropy principle</u>:

"In the problem of prediction, the maximization of entropy is not an application of a law of physics, but merely a method which ensures that no unconscious arbitrary assumptions have been introduced."

I: physics, prior info

signal space: $s \in \mathbb{R}^{\infty}$ data space: $d \in \mathbb{R}^{|d| \ll \infty}$




$P(s|d,I)$ more concentrated than $P(s|I)$

$\Rightarrow$ information gain

Bayes theorem:

measurement likelihood, prior knowledge

$$P(s|d,I) = \frac{P(d,s|I)}{P(d|I)} = \frac{P(d|s,I)\,P(s|I)}{\int \mathcal{D}s\, P(d,s)}$$

normalization

Maximum Ignorance/Entropy: only $1^{st}, 2^{nd}$ moments known $\Rightarrow$ Gaussian PDF

$$P(s\,|\,\underbrace{m_x = \langle s_x \rangle,\ S_{xy} = \langle s_x s_y \rangle}_{I}) = \mathcal{G}(s-m, S)$$

$$= \frac{1}{|2\pi S|^{1/2}} \exp\left[-\frac{1}{2}\sum_{xy} (s_x - m_x)\, S^{-1}_{xy}\, (s_y - m_y)\right]$$

$\Rightarrow P(s|I)$ prior.

linear measurement: $d_i = \sum_x R_{ix}\, s_x + n_i$

noise uncorrelated with signal, zero mean, known variance: $P(n|I) = \mathcal{G}(n, N)$

$$P(d|s, I) = \mathcal{G}(\underbrace{d - Rs}_{=n}, N) \quad \text{likelihood}$$

goal: minimal least squared error reconstruction

error: $\varepsilon_x^2 = \langle (s_x - m_x)^2 \rangle$ minimal for $\forall x$ $\quad \langle \ldots \rangle \equiv \int \mathcal{D}s \int \mathcal{D}d\, P(s,d) \ldots$

optimal linear filter: $\boxed{m_x = \sum_i F_{xi}\, d_i \quad , \quad m = F d}$

$$\frac{\partial \varepsilon_x}{\partial F_{xi}} = \frac{\partial}{\partial F_{xi}} \left\langle \left(s_x - \sum_j F_{xj} d_j\right)^2 \right\rangle = -2 \left\langle \left(s_x - \sum_j F_{xj} d_j\right) d_i \right\rangle = 0$$

$$\Rightarrow \langle s_x d_i \rangle = \sum_j F_{xj} \langle d_j d_i \rangle \Rightarrow \langle s d^t \rangle = F \langle d d^t \rangle \Rightarrow \boxed{F = \langle s d^t \rangle \langle d d^t \rangle^{-1}}$$

$$\boxed{m = \langle s d^t \rangle \langle d d^{\cancel{t}} \rangle^{-1} d}$$

$$d = Rs + n \;\Rightarrow\; \langle s d^t \rangle = \langle s (Rs+n)^t \rangle = \underbrace{\langle s s^t \rangle}_{S} R^t + \underbrace{\langle s n^t \rangle}_{0} = S R^t$$

$$\langle d d^t \rangle = \langle (Rs+n)(Rs+n)^t \rangle = R \underbrace{\langle s s^t \rangle}_{S} R^t + R \underbrace{\langle s n^t \rangle}_{0} + \underbrace{\langle n s^t \rangle}_{0} R^t + \underbrace{\langle n n^t \rangle}_{N}$$

$$= R S R + N$$

$$\boxed{m = S R^t (R S R + N)^{-1} d}$$ Wiener filter

interferometer measures in Fourier domain: $\vec{k} = \vec{b}/\lambda$ (baseline), $R_{kx} = e^{i\vec{k}\cdot\vec{x}}$ also complex conjugation!

stat. translational invariant sky: $S_{kk'} = \sum_{xy} e^{ikx} S_{xy} e^{-ik'y} = (R S R^{\dagger})_{kk'} = \delta(k-k') P_s(k)$

noise covariance $N_{kk'} = \delta(k-k') P_n(k)$

Fourier image: $m_k = \sum_x e^{ikx} m_x = (Rm)_k = (R S R^{\dagger} (R S R + N)^{-1} d)_k$

$$= P_s(k) / (P_s(k) + P_n(k))\, d_k = \frac{d_k}{1 + (P_n/P_s)(k)}$$

$$= d_k \begin{cases} 1 & \text{for } P_s \gg P_n \leftarrow \text{„uniform" weighting} \\ (P_s/P_n)(k) & \text{for } P_n \ll P_s \leftarrow \text{„natural" weighting} \end{cases}$$

| signal prior | | optimal algorithm |
|---|---|---|
| Gaussian signal statistics | $\longrightarrow$ | Wiener filter, robust weighting, |
| signal sparse in pixel space (point sources) | $\longrightarrow$ | CLEAN |
| signal sparse in Wavelet space | $\longrightarrow$ | Compressed sensing |
| other statistics | $\longrightarrow$ | tailor algo with information field theory |
| unknown statistics | $\longrightarrow$ | critical filter methodology |
| hidden prior (can be reverse ungeneered) | $\longleftarrow$ | some algorithm |

Since the finite data samples the signal (with infinite many d.o.f.) coarsely, incompletely, and with noise, always a prior or regularization has to be present ~~But~~ Not all ways are the advocates of a method aware of this.