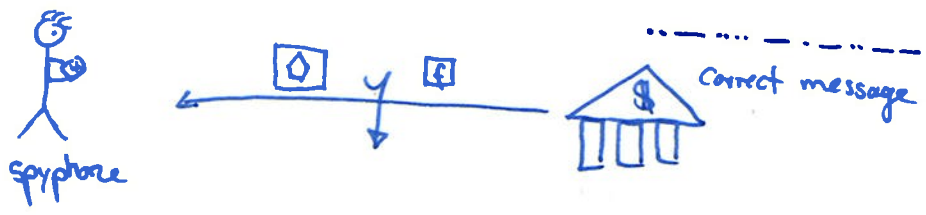
# Lecture 2 : Inference

11th of Feb 2019

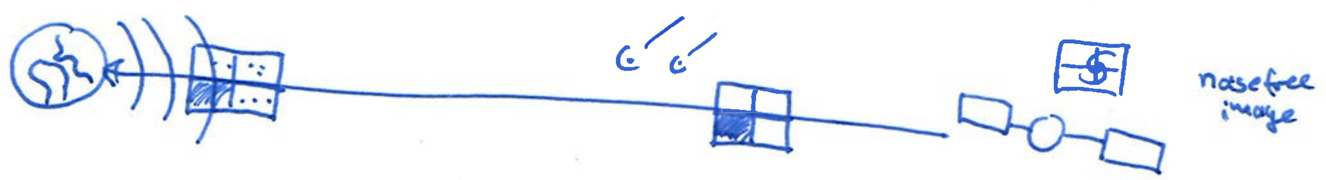
- Inference: what can possibly go wrong? Let's investigate this by a decryption example.

- Setup :

Either:




Or: Noisy downlink




present image by data matrix $M$ = pxp matrix ; $M_{ij}$ are pixels of image

- $\vec{d} = \text{vec}[M] \Rightarrow M_{ij} = d_{i\cdot p+j}$

- a (linear) redundant encryption could be
$E\,\vec{d} = \vec{e}$ with $\dim(\vec{e}) > \dim(\vec{d})$

- for example could have

$$\vec{d} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} \quad \text{and} \quad E = \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \\ 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{pmatrix} \Rightarrow E\,\vec{d} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} = \vec{e}$$

- We could decrypt this with

$$D\,\vec{e} \overset{!}{=} \vec{d} \Rightarrow D = \frac{1}{2}\begin{pmatrix} 1&0&0&0&1&0&0&0 \\ 0&1&0&0&0&1&0&0 \\ 0&0&1&0&0&0&1&0 \\ 0&0&0&1&0&0&0&1 \end{pmatrix}$$

Advantage: · $\vec{d}+\vec{n}$ extremely noisy.

· but $\vec{e}+\vec{n}$ after decryption less noisy (but random!)

$D(\vec{e}+\vec{n}) = \hat{\vec{d}}$ hat: it is now random (an „estimator")
since $\vec{n}$ is random

• Bank example: bank uses D, but you (for some reason) use D'!
⇝ you'll decode the wrong message

## * Noisy encryption

· true parameters of „Nature": $\vec{\Theta} = (\Theta_1, \Theta_2, \vec{\Theta}_3, \dots, \Theta_n)$
e.g. true neutrino masses
or true number of exoplanets orbiting a star

· then astronomical data „encrypt" these true values with a noisy operator:

$$\boxed{\vec{d} = \hat{E}\,\vec{\Theta}}$$

hat: it is noisy // random (boldfont in the script)

· Clearly, the „message" is still there (the true values), but now encoded by noise.

· „Inference": decode the message, i.e. invert $\hat{E}$ as good as you can.

· hope: $\dim(\vec{d}) \gg \dim(\vec{\Theta}) \Rightarrow \hat{E}$ a very redundant encryption

· Now: $\vec{\Theta}$: true parameters of Nature

decryption: $\hat{E}$ random, hence decryption operator $\hat{D}$ must be random.

$\Rightarrow \underbrace{\hat{D}}_{\text{random}}\,\underbrace{\vec{d}}_{\text{random}} = \hat{\vec{\Theta}}$ ⇒ inferred parameters now also random,
where randomness is called uncertainty.

· If $\hat{D}$ overly approximated, then $\hat{\vec{\Theta}}$ far off $\vec{\Theta}_t$ : „bias"

# From physics (single truth) to data

• Example 1) Stars:

Stefan-Boltzmann law relates stellar luminosities $L$ to stellar areas $A$ and temperatures $T$ via

$$L = \sigma_T A T^4$$

• Let's generate a stellar population by drawing

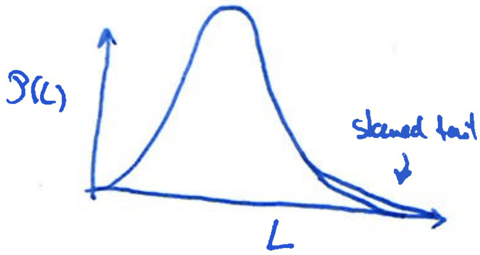
$$T \sim \mathcal{G}(T_0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(T-T_0)^2}{\sigma^2}\right)$$

⇝ which distribution $\rho(L)$ ensues?

• $\rho(L)\,dL = \rho(T)\,dT \quad \Rightarrow \quad \rho(L) = \rho(T(L)) \left|\frac{dT}{dL}\right|$

$$T(L) = \sqrt[4]{\frac{L}{\sigma_T A}} \quad \Rightarrow \quad \frac{dT}{dL} = \frac{1}{4}\left(\frac{L}{\sigma_T A}\right)^{-\frac{3}{4}}$$

$$\Rightarrow \quad \rho(L) = \mathcal{N} \exp\left(-\frac{1}{2} \frac{\left(\sqrt[4]{\frac{L}{\sigma_T A}} - T_0\right)^2}{\sigma^2}\right) \cdot \underbrace{\frac{1}{4}\left(\frac{L}{\sigma_T A}\right)^{-3/4}}_{\text{produces skewness}}$$



Now $T$ with intrinsic scatter as before: $T \sim \mathcal{G}(T_0, \sigma^2)$

and $A$ with intrinsic variation: $A \sim \mathcal{P}(A)$

- How does $L$ now scatter?

$\rightsquigarrow$ $\mathcal{P}(L,A,T) = \mathcal{P}(L|A,T)\,\mathcal{P}(A,T)$ [conditional]

$\rightsquigarrow$ know: $\mathcal{P}(L|A,T) = \delta_D(L - \sigma_T A T^4)$ [physics!]

$\rightsquigarrow$ and according to my setup: $\mathcal{P}(A,T) = \mathcal{P}(A)\,\mathcal{P}(T)$ [independence]

$$\Rightarrow \boxed{\mathcal{P}(L,A,T) = \delta_D(L - \sigma_T A T^4)\,\mathcal{P}(A)\,\mathcal{P}(T)}$$

we want: $\mathcal{P}(L) = \int_A \int_T \mathcal{P}(L,A,T)\,dA\,dT$ [marginalization]

$$= \int_A \int_T \mathcal{P}(A)\,\mathcal{P}(T)\,\underbrace{\delta_D(L - \sigma_T A T^4)\,dA\,dT}$$

This we already solved!
$\mathcal{P}(L) = \int \delta_D(L - \sigma_T A T^4)\,\mathcal{P}(T)\,dT$
$= \mathcal{P}[T(L)]\,\left|\frac{dT}{dL}\right|$

killed T integral

$$= \int_A \mathcal{P}(A)\,\left|\frac{dT}{dL}\right|\,\mathcal{P}[T(L)]\,dA$$

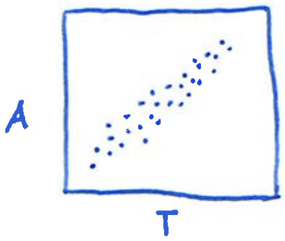
$$= \int_A \mathcal{P}(A)\cdot \mathcal{N}\left|\frac{1}{4}\left(\frac{A\sigma_T}{L}\right)^{+\frac{3}{4}}\right| \exp\left(-\frac{1}{2\sigma^2}\left[\left(\frac{\sigma_T A}{L}\right)^{-1/4} - T_0\right]^2\right) dA$$

- For example, could now put

$$\mathcal{P}(A) = \mathrm{Gamma}(\nu, A_0) = \frac{A_0^{\nu} A^{\nu-1} \exp(-A_0/A)}{\Gamma(\nu)}$$

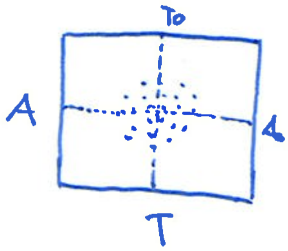
$$\boxed{\text{Then } \mathcal{P}(L) = \int_A \mathcal{N}\,\frac{A_0^{\nu} A^{\nu-1} \exp(-A_0/A)}{\Gamma(\nu)}\left|\frac{1}{4}\left(\frac{A\sigma_T}{L}\right)^{+3/4}\right| \exp\left(-\frac{1}{2\sigma^2}\left[\left(\frac{\sigma_T A}{L}\right)^{-1/4} - T_0\right]^2\right) dA}$$

Should have an analytical solution, since exp and powerlaws actually, I want to measure $\sigma_T$ from noisy measurements of $L$, from a population of stars with scatter in $T$ and $A$:

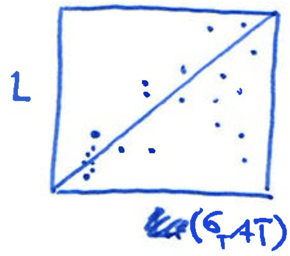



$\Rightarrow$ violates assumption $\mathcal{P}(A,T) = \mathcal{P}(A)\,\mathcal{P}(T)$

mathematical plot representing the ESO image in the slides

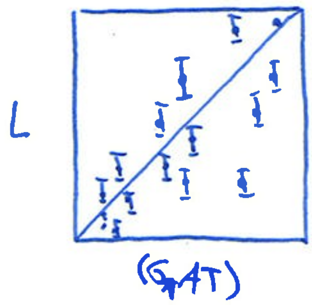



$\Rightarrow$




stellar variability

$\Rightarrow$ now additionally with measurement noise




$\leftarrow$ mathematical plot representing the noisy ESO image in the slides

- have $\hat{L} \sim \mathcal{P}(\hat{L} \mid L) \underset{\text{e.g.}}{=} \frac{1}{\sqrt{2\pi\sigma_L^2}} \exp\left(-\frac{1}{2\sigma_L^2}(\hat{L}-L)^2\right)$

because detectors tend to behave this way ant: $P(\sigma_T | \hat{L})$

Have: $$P(\sigma_T | \hat{L}) = \int P(\sigma_T, A, T, L | \hat{L}) \, dA \, dT \, dL \qquad \text{[marginal]}$$

$$= \int \frac{P(\sigma_T, A, T, L, \hat{L})}{P(\hat{L})} \, dA \, dT \, dL$$

independent

$$= \int \frac{P(\sigma_T, A, T \,|\, L, \cancel{\hat{L}}) \; P(L, \hat{L})}{P(\hat{L})} \, dA \, dT \, dL$$

$$= \int P(\sigma_T, A, T | L) \; P(\hat{L} | L) \; \frac{P(L)}{P(\hat{L})} \, dA \, dT \, dL$$

$$= \int P(\sigma_T, A, T, L) \; \frac{\overset{\text{Gauss}}{P(\hat{L} | L)}}{P(\hat{L})} \, dA \, dT \, dL$$

$$= \int \underbrace{P(L | A, T, \sigma_T)}_{\delta_D:\ \text{physics!}} \; \underbrace{P(A, T, \cancel{\sigma_T})}_{\text{factorizes}} \; \frac{G(\hat{L} | L)}{P(\hat{L})} \, dA \, dT \, dL$$

$$= \int \underline{\delta_D}(L - \sigma_T A T^4) \; P(A) \, P(T) \; \frac{G(\hat{L} | L)}{P(\hat{L})} \, dA \, dT \, \underline{\underline{dL}}$$

$$= \int \frac{\overset{\text{Gamma}}{P(A)} \, \overset{\text{Gauss}}{P(T)} \; G(\hat{L} \,|\, \sigma_T A T^4)}{P(\hat{L})} \, dA \, dT \, .$$

$$= P(\sigma_T | \hat{L})$$