Source separation.

Problems, principles of solutions, applications

Christian JUTTEN

GIPSA-lab, Dept. Images & Signal

Grenoble

Christian.Jutten@inpg.fr

Summary

- 1. The problem of source separation
- 2. Dependence measures
- 3. Linear instantaneous mixtures
- 4. Algorithms
- 5. Convolutive mixtures
- 6. Nonlinear mixtures
- 7. Semi-blind approaches and algorithms
- 8. Applications

1. The problem of source separation

Source separation

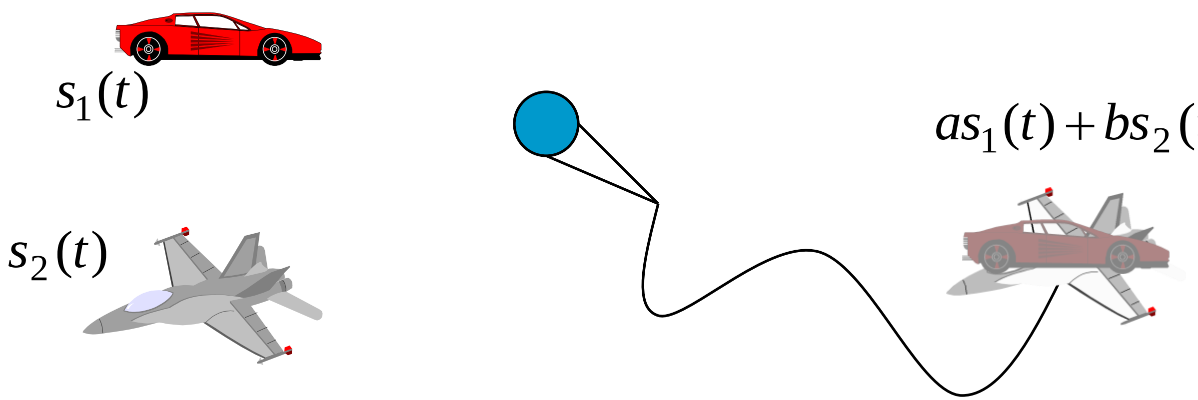
Motion decoding

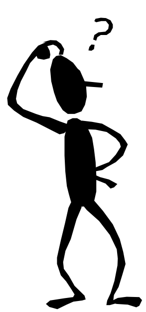
Blind source separation and ICA

Undeterminacies

1.1. Source separation: the problem

- The signal received by a sensor (electrode, antenna, microphone, etc.) is an intricated signal.

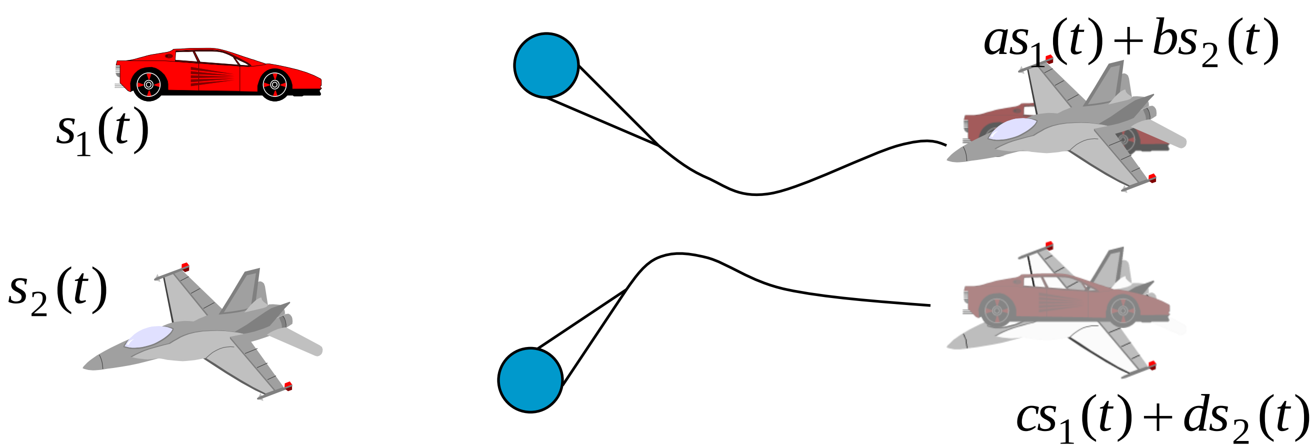




Is it possible to retrieve the different signals (sources) from the mixture ? If yes, how ?

Source separation: main idea

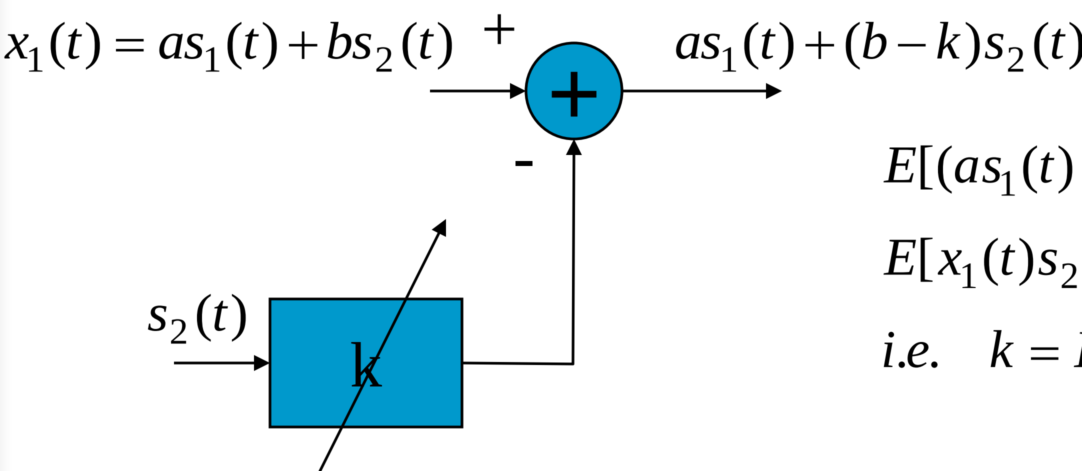
- It is possible using a few observations
  - more sensors than sources
  - different mixtures (observations) $\left(\frac{a}{b} \neq \frac{c}{d}\right)$



- Spatial **diversity**

Source separation with priors

- With priors on sources
  - different frequency ranges: classical simple filtering,
  - one reference (Widrow-Hoff method): adjust *k* so that the error is decorrelated from the reference



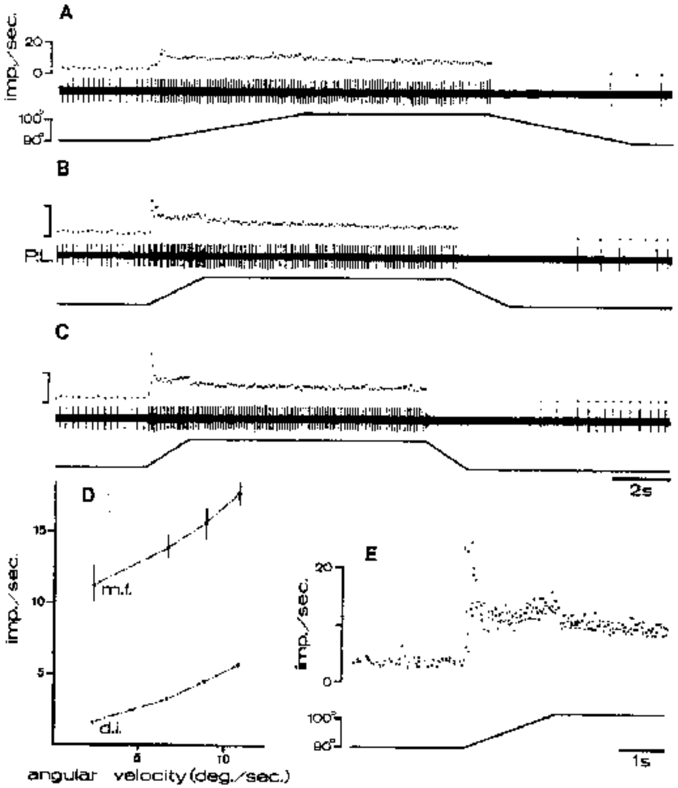
$$E[(as_1(t) + (b-k)s_2(t))s_2(t)] = 0$$

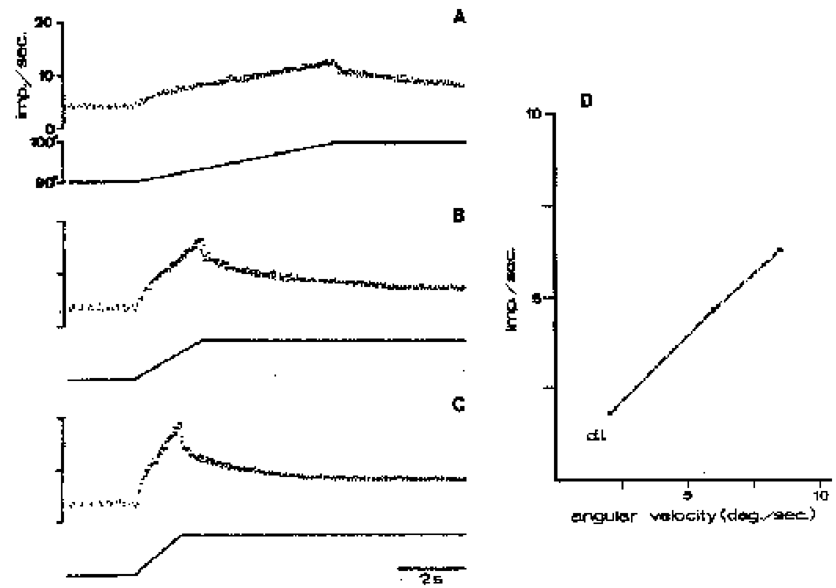
$$E[x_1(t)s_2(t)] - kE[s_2^2(t)] = 0$$

$$i.e. \quad k = E[x_1(t)s_2(t)] / E[s_2^2(t)]$$

- But if sources have same frequency ranges, and if there is no reference ?

1.2. Origin: Motion decoding



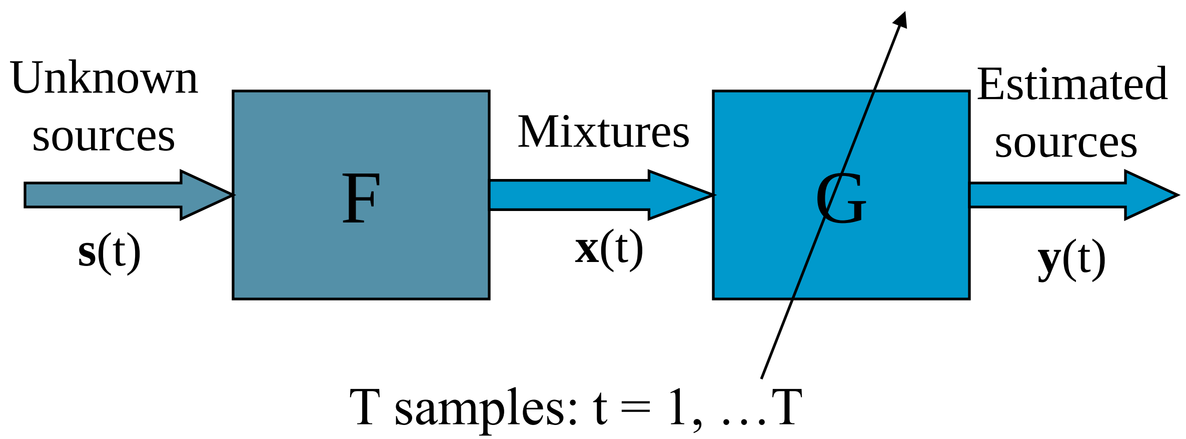


Linear model
$$\begin{cases} f_I(t) = a_{11}\,p(t) + a_{12}\,v(t) \\ f_{II}(t) = a_{21}\,p(t) + a_{22}\,v(t) \end{cases}$$

Can one retrieve p(t) et v(t) from the mixtures $f_I(t)$ and $f_{II}(t)$?
There are some biological evidences…
But, under what conditions ? And how ?

1.3. BSS problem

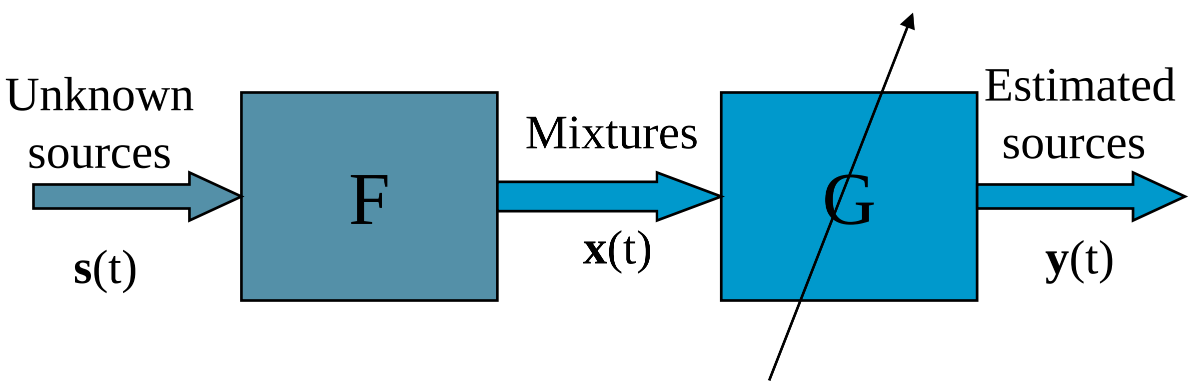
- Assumption on the mixture model F
  - linear instantaneous (without memory) model,
  - linear convolutive (with memory) model,
  - nonlinear model.
- One chooses a separation model G, suited to the mixture model



ICA method

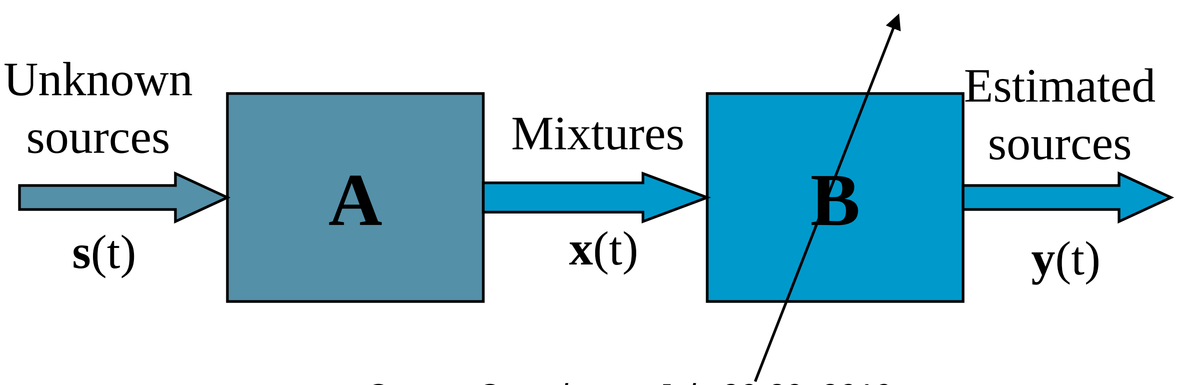
- Assume the model F is known, and G is suited to F

- Without priors, the problem is impossible
- A possible source prior: statistical independence
  - Independent Component Analysis: ICA method,
  - idea: estimate G so that the Y_i's become independent
  - since prior is very weak, the problem is called ***blind***



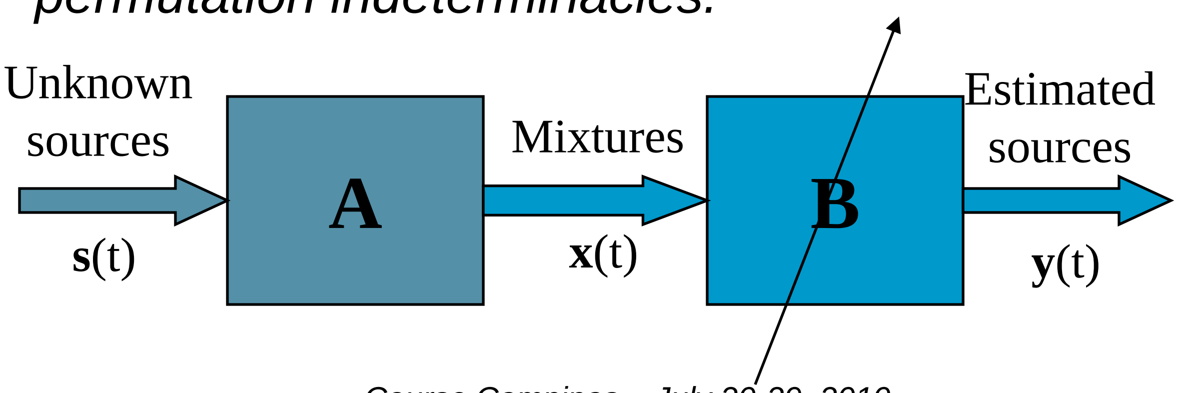
Linear case: can BSS be solved ?

- Impossible (Darmois 1953) for
  - *independent identically distributed (iid) Gaussian sources*
- Possible for
  - *iid **non Gaussian** sources*
  - *Gaussian **non iid** sources, e.g. non temporally independent (colored), or non identically distributed (statistics is changing: nonstationarity)*



ICA: linear instant. mixtures

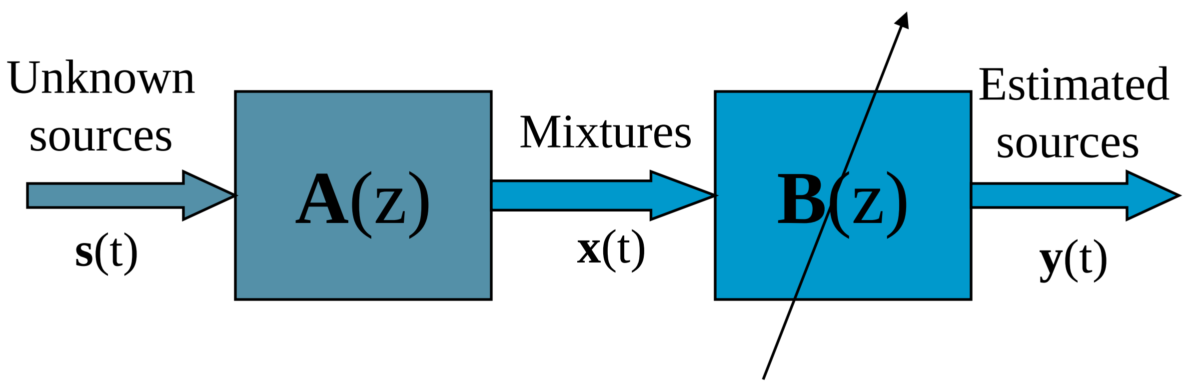
- Question : Y_i independent, $\forall i \Leftrightarrow \quad Y \approx S$?
- Theoretical results (Comon, Signal Processing 1994)
  - Let be **x**(t) = **A s**(t) a linear instantaneous regular mixture (**A** is a regular matrix) of independent sources **s**(t), whose at most one is Gaussian, signals **y**(t) = **B x**(t) are independent iff **BA** = **DP**
  - *Independence is equivalent to separation up to scale and permutation indeterminacies.*



ICA: linear convol. mixtures

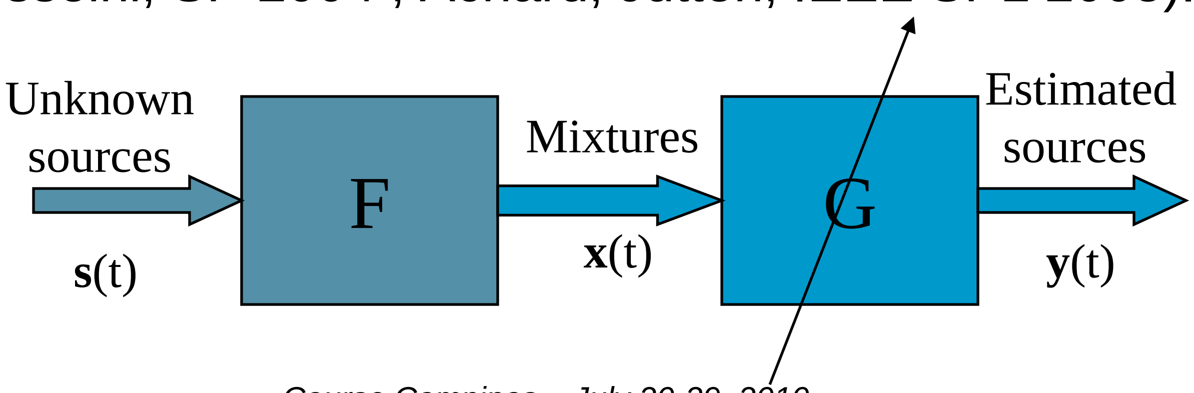
- Theoretical results for convolutive mixtures (Weinstein, Yellin, IEEE SP, 94 ; Nguyen, Jutten, SP, 1995).
  - Let be **x**(t) a linear convolutive mixture (**A**(z) is an invertible filter matrix) of independent sources, whose at most one is Gaussian, signals **y**(t)=[**B**(z)]**x**(t) are independent iff **B**(z)**A**(z) = **D**(z)**P**.

Independence is equivalent to separation up to filter and permutation undeterminacies.



ICA: nonlinear mixtures

- Theoretical results for nonlinear mixtures
  - For general nonlinear mixtures, Y independence does not insure source separation (Darmois, 1953 ; Hyvärinen, Pajunen,1998).
  - Particular nonlinear mixtures are separable, e.g. the post-nonlinear models (Taleb, Jutten, IEEE SP 1999 ; Babaie-Zadeh, Jutten, Eusipco 2002 ; Jutten, Babaie-Zadeh, Hosseini, SP 2004 ; Achard, Jutten, IEEE SPL 2005).



ICA: under- or over-determined

- More mixtures than sources: over-determined mixtures
  - there is solution provided than the mixture is invertible
- More sources than mixtures:under-determined mixtures
  - there is no solution,
  - even, if the mixture has been estimated or is known, since it is not invertible, sources cannot be deduced: “identification of F” and “restoration of sources” are two different problems,
  - for linear mixtures, one can show (Taleb, Jutten, ICASSP 99) that the sources are known up to a random vector !
  - Solution if extra priors, e.g. discrete or sparse sources

ICA: linear noisy mixtures

- The model is:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{b}(t)$$

  where b(t) is assumed non correlated with s(t).

- Basically, there are two problems:
  - noise (except if it is Gaussian, and ICA use higher-order statistics) implies a biased estimation of **B** = **A**$^{-1}$
  - even if **B** is perfectly estimated, i.e. **B** = **A**$^{-1}$, the estimated sources are:

$$\mathbf{y}(t) = \mathbf{s}(t) + \mathbf{A}^{-1}\mathbf{b}(t)$$

- The noise can be amplified.

1.4. Understand the undeterminacies

- Scale and permutation undeterminacies can be understood according to various ways.
- For a linear model, the sensor *i* receives the mixture :

$$
\begin{aligned}
x_i(t) &= a_{i1}s_1(t) + a_{i2}s_2(t) + \ldots + a_{in}s_n(t) \\
&= \sum_j a_{ij} s_j(t) \\
&= \sum_j \left(\frac{a_{ij}}{\alpha_j}\right)\left(\alpha_j s_j(t)\right)
\end{aligned}
$$

- With independence: if **y**(t) has independent components, then **DPy**(t) too.

Consequence of undeterminacies

- Undeterminacies mean that we cannot estimate the power of the sources (in linear mixtures)
- As a consequence,
  - we cannot estimate n parameters among the $n \times n$ of the mixing matrix,
  - we can then, without loss of generality, impose the main diagonal of **A** to be equal to 1
  - we also can parameterize the separating matrix **B** with fixing n parameters, or with column normalization
  - we can arbitrarily fix the power of estimated sources,

2. Dependence measures

Independence definition

Dependence measures

2.1. Supervised or unsupervised

- ICA is based on source independence
  - Independence (or dependence) measure does not need external information, i.e. no supervisor
  - ICA is then basically an unsupervised estimation method
  - it is also called “blind” like in similar deconvolution or equalization problems

- Beyond ICA, one can look for using priors

- In this Section, one focuses on independence measure for ICA

2.1. Independence definitions

- With words…
- From probability
  - for 2 random variables:

$$p_{Y_1 Y_2}(u_1, u_2) = p_{Y_1}(u_1)\, p_{Y_2}(u_2)$$

  - for *N* random variables:

$$p_{Y_1 Y_2 \ldots Y_N}(\mathbf{u}) = p_{Y_1}(u_1)\, p_{Y_2}(u_2) \ldots p_{Y_N}(u_N)$$

$$= \prod_{k=1}^{N} p_{Y_k}(u_k)$$

- Problem: equality of multivariate functions

Independence definitions

- First characteristic functions
  - it is the inverse Fourier transform of the probability density function:

$$\varphi(\nu) = E[\exp(j\nu u)] = \int p(u)\exp(j\nu u)du$$

$$\varphi(\mathbf{v}) = E[\exp(j\mathbf{v}^T\mathbf{u})] = \int \cdots \int p(\mathbf{u})\exp(j\mathbf{v}^T\mathbf{u})\mathbf{du}$$

  - as a consequence, there is the same information in the pdf and in the characteristic function.
- Second characteristic function
  - it is the log of the first characteristic function:

$$\phi(\nu) = \log\varphi(\nu)$$

$$\phi(\mathbf{v}) = \log\varphi(\mathbf{v})$$

Independence definitions

- From the definition (random variable):

$$\varphi(\nu) = E[\exp(j\nu u)] = \int p(u)\exp(j\nu u)du$$

1 for ν = 0

$$\phi(\nu) = \log\varphi(\nu)$$

1 for ν = 0

- The SCF is equal to 0 for ν = 0.

- One can show the SCF is continuous near 0. Then, it can be expanded in Taylor series.

Independence definitions

- Cumulant definition
  - it is the coefficients of the expansion of the SCF:

$$\phi(\nu) = \log \varphi(\nu) = \sum_{p=1}^{+\infty} \kappa_p \frac{(j\nu)^p}{p!}$$

- In other words, the order-p cumulant is:

$$\kappa_p = (-j)^p \left. \frac{d^p \phi(\nu)}{d\nu^p} \right|_{\nu=0}$$

Independence definitions

- Exercice: compute the 2nd, 3rd and 4th order cumulants of a zero-mean variable

$$\phi(\nu) = \log \varphi(\nu) = \sum_{p=1}^{+\infty} \kappa_p \frac{(j\nu)^p}{p!}$$

- Write the Taylor expansion:

$$\ln \varphi(\nu) = \ln \varphi(0) + \left.\frac{d \ln \varphi(\nu)}{d\nu}\right|_{\nu=0} \nu + \left.\frac{d^2 \ln \varphi(\nu)}{d\nu^2}\right|_{\nu=0} \frac{\nu^2}{2!} + \ldots$$

and identify to the above expansion.

Independence definitions

- Independence and characteristic functions

$$p_{Y_1Y_2\ldots Y_N}(\mathbf{u}) = p_{Y_1}(u_1)\,p_{Y_2}(u_2)\ldots p_{Y_N}(u_N)$$

$$\varphi(\mathbf{v}) = \varphi(v_1)\varphi(v_2)\cdots\varphi(v_N)$$

$$\log\varphi(\mathbf{v}) = \log\varphi(v_1) + \log\varphi(v_2) + \cdots + \log\varphi(v_N)$$

$$\phi(\mathbf{v}) = \phi(v_1) + \phi(v_2) + \cdots + \phi(v_N)$$

- Independence = cancellation of all cross-cumulants

  considering Taylor expansion near 0, since right-side term does not content cross-terms, contrary to the left-side term

- Problem: there are an infinite number of cumulants !

Independence definitions

- *p*-order cumulants are function of statistical moments up to order *p*
- The first cumulants for a zero mean variable are:

$$\kappa_1 = \mu_1 = 0$$
$$\kappa_2 = \mu_2$$
$$\kappa_3 = \mu_3$$
$$\kappa_4 = \mu_4 - 3\mu_2$$
$$\ldots$$

where $\mu_p = E[X^p]$ is the *p*-order moment.

2.2. Dependence measures

- Kullback-Leibler divergence

$$KL(f \| g) = \int \ldots \int f(\mathbf{u}) \log \frac{f(\mathbf{u})}{g(\mathbf{u})} \mathbf{du}$$

- It is a real number which measures the divergence between two distributions
- KL divergence is positive and vanishes if and only if $f = g$ (this is easy to show using Jensen inequality)

2.2. Dependence measures

- Kullback-Leibler divergence between the joint pdf and the product of marginal pdfs of the random vector **Y**

$$KL(p_{\mathbf{Y}} \| \prod p_{Y_i}) = \int \ldots \int p_{\mathbf{Y}}(\mathbf{u}) \log \frac{p_{\mathbf{Y}}(\mathbf{u})}{\prod_i p_{Y_i}(u_i)} \mathbf{du}$$

- The above divergence is positive and vanishes if and only if **Y** has independent components, i.e.:

$$p_{Y_1 Y_2 \ldots Y_N}(u,v) = p_{Y_1}(u) p_{Y_2}(v) \ldots p_{Y_N}(v)$$

$$= \prod_{k=1}^{N} p_{Y_k}(v)$$

- This KL divergence is equal to Mutual Information I(**Y**)

Dependence measures

- Mutual information (MI)

$$I(\mathbf{Y}) = \int p_{\mathbf{Y}}(\mathbf{u}) \log \frac{p_{\mathbf{Y}}(\mathbf{u})}{\prod_i p_{Y_i}(u_i)} \mathbf{du}$$

$$= \sum_{i=1}^{N} H(Y_i) - H(\mathbf{Y})$$

where the marginal and joint differential entropies are defined as:

$$H(Y_i) = -\int p_{Y_i}(u) \log p_{Y_i}(u) du$$

$$H(\mathbf{Y}) = -\int \cdots \int p_{\mathbf{Y}}(\mathbf{u}) \log p_{\mathbf{Y}}(\mathbf{u}) \mathbf{du}$$

Dependence measures

- Mutual information (MI)

$$I(\mathbf{Y}) = \sum_{i=1}^{N} H(Y_i) - H(\mathbf{Y})$$

$$H(Y_i) = -\int p_{Y_i}(u) \log p_{Y_i}(u)\, du, \quad H(\mathbf{Y}) = -\int \cdots \int p_{\mathbf{Y}}(\mathbf{u}) \log p_{\mathbf{Y}}(\mathbf{u})\, \mathbf{du}$$

- Problem: estimation of MI requires estimation of **joint** and **marginal** probability density functions

Simpler dependence measures ?

- Decorrelation = second order independence

$$\text{decorrelation} \not\Rightarrow \text{independence}$$
$$\text{independence} \Rightarrow \text{decorrelation}$$

- We will see later that, using algebraic arguments, decorrelation is not sufficient

- However, decorrelation is a first step toward independence...