# Data analysis
## spatio-temporal data and hierarchical models

Foundation (Ch2)

Djemel Ziou

# Probability density function

- A random variable (rv) **x** follow a probability law L means that the set of values $\{x_1, \ldots x_n\}$ can be sampled from L.
- Their exists several probability distributions such as normal, Poisson, beta, Dirichlet, gamma, binomial, and multinomial. Some are for discrete random variables and others for continuous variables. For example,
  - the rv whose values are the number of people over a period of time arriving at the university is discreet (often considered as a Poisson law).
  - the rv whose values are the price of items in a store is continuous.
- The sum of n rv $\mathbf{x_1}, \ldots \mathbf{x_n}$ independent and identically distributed (iid) of a finite expectation μ and a variance $\sigma^2$ converges (in probability) to Gaussian random variable of mean nμ and variance $n\sigma^2$.

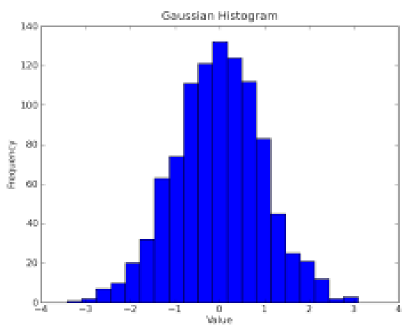
# Probability density function -example

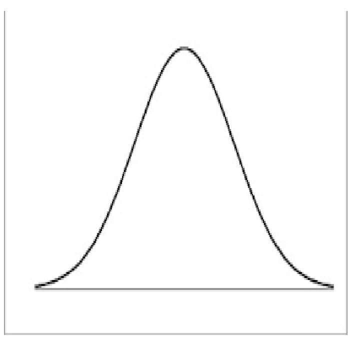
- Probability density function (pdf) of normal random variable (rv)
  - 1D

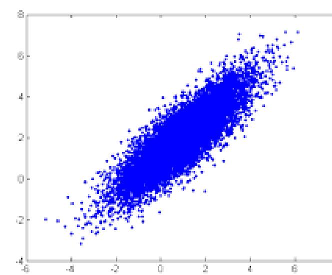
$$[x] = \exp(-(x-\mu)^2 / 2\sigma^2) / \sqrt{2\pi}\sigma$$

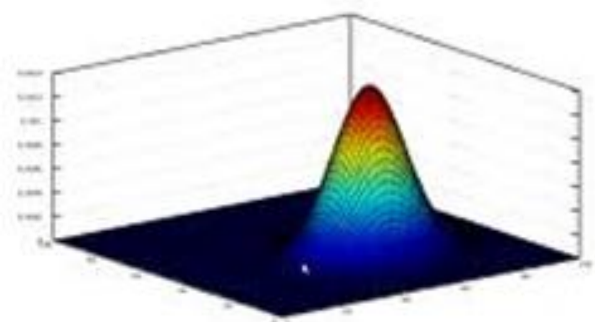
  - mD

$$[\vec{x}] = \exp(-(\vec{x}-\vec{\mu})^t \Sigma^{-1} (\vec{x}-\vec{\mu}) / 2) / (2\pi)^{-d/2} |\Sigma|^{1/2}$$







# Probability density function – statistics

- Let x be a rv and [x] its pdf (a probability mass in the discrete case). The expectation is
  - continuous 1D $$E(x) = \int_{\Omega} x[x]dx$$
  - discrete 1D $$E(x) \approx \sum_{x} x[x]$$
- Variance
  - 1D $$V(x) = E(x^2) - E^2(x)$$
  - mD

  Let x and y be random vectors of dimension d with know joint pdf. The covariance

$$\mathrm{cov}(\vec{x}, \vec{y}) = E(\vec{x} - E(\vec{x}))E(\vec{y} - E(\vec{y}))^t$$

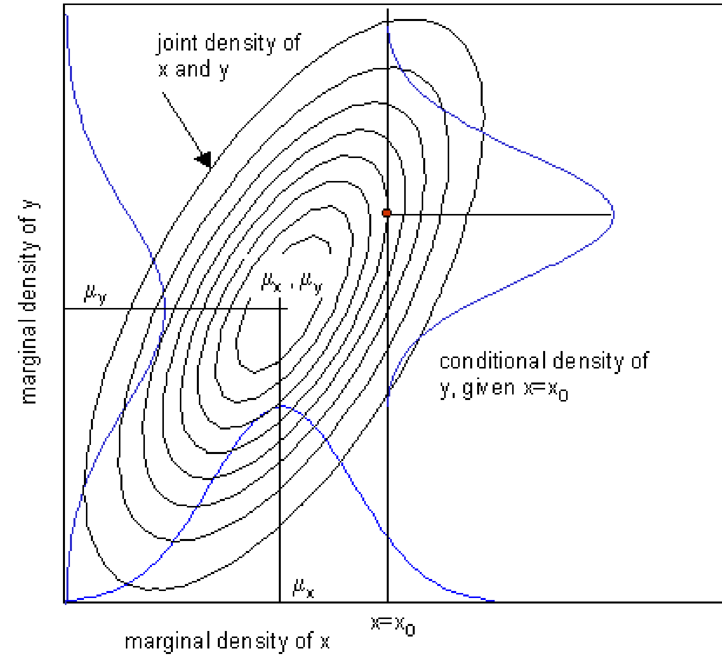
where $$E(\vec{v}) = (E(v_1), \cdots, E(v_d))^t$$

# Probability density function - inference

- The uncertainty in the data, the process, and the parameters may leads to an uncertain conclusion.
  - example, if you throw a coin 7 times, can we conclude if it is fake or not?
- Inference means estimation of parameters or probability density functions (probability mass).
  - example, let a random variable x and $X = \{x_1 \ldots x_n\}$ data sampled from a 1D normal pdf. Find the estimator of the mean and variance.
  - perform the same reasoning by assuming that the data is sampled from a mixture of normal pdfs.

# Probability density function - use

- The pdf can be used for
  - estimating parameters of models (e.g. time series)
  - fit the data (e.g. the histogram fit by the Gaussian)
- For this, we use
  - joint pdf to model events.
  - marginal pdf to ignore events.
  - conditional pdf to model causality.
- Let x and y two random variables, we have




$$[x,y]=[x|y][y]=[y|x][x]$$
$$[x]=\int [x,y]dy=\int [x|y][y]dy=\int [y|x][x]dx$$
$$[x|y]=[y|x][x]/\int [x|y][y]dy$$

# Probability density function - use

- Example: It is proposed to classify ($w_0$, $w_1$) fish from a photo. Consider two features: the gray level $x_1$ and the length $x_2$. The length can not be estimated because of occlusions (hidden variables).

- So we must calculate

$[w_1|x_1]=[w_1,x_1]/[x_1]=\int_\Omega [w_1,x_1,x_2]/[x_1]dx_2=$
$\int_\Omega[w_1|x_1,x_2][x_1,x_2]/[x_1]dx_2$. If $x_1$ and $x_2$ are independent, then
$[w_1|x_1]=\int_\Omega[w_1|x_1,x_2][x_2]dx_2$.

# Probability density function - use

One can use the gamma pdf

$[x_2] = x_2^{a-1} b^a e^{-bx_2} / GAMMA(a)$

$[w_1 | x_1, x_2] = \exp(a_1 x_1 + a_2 x_2 + b_1)/(1+\exp(a_1 x_1 + a_2 x_2 + b_1)$.

- For the classification, we can use the Bayes decision rule

$[w_1 | x_1] > [w_0 | x_1]$ then $x_1$ belongs to the class $w_1$ and to $w_0$ otherwise

# Bayes: Prior, likelihood, posterior

- Suppose that the distribution of the height of women $(F)$ in Shenzhen is normal with the mean 1.65 m and 0.16 m standard deviation (SD):

$$p(t|F) = \frac{1}{0.16\sqrt{2\pi}} \exp(-\frac{1}{2}(\frac{t-1.65}{0.16})^2)$$

- And that of men $(H)$ is also normal of mean 1.75 m and SD 0.15 m:

$$p(t|H) = \frac{1}{0.15\sqrt{2\pi}} \exp(-\frac{1}{2}(\frac{t-1.75}{0.15})^2)$$

- We would like to decide about the gender of a person who measures 1.60 m

$$p(1.60|F) \approx 2.37 \qquad p(1.60|H) = 1.61$$

The values are greater than 1?

# Bayes: Prior, likelihood, posterior

- Now, suppose that there is as many men as women in the Shenzhen, what is the probability that the person is a woman?
- Bayes' formula provides the answer

$$p(F|1.60) = \frac{0.5\,p(1.60|F)}{0.5\,p(1.6|F) + 0.5\,p(1.60|H)} \approx 60\%$$

- And, $p(H|1.60) = 40\%$
- The person is assigned to group $F$ with 60% and to H with 40%
- Again, suppose that a person is selected at random in a stadium during a soccer game where 30% are women and 70% men. Thus,

$$p(F|1.60) = 39\% \qquad et \qquad p(H|1.60) = 61\%$$

# Bayes: Prior, likelihood, posterior

- Let the data D={$x_1 \ldots x_n$} sampled from a pdf with a parameter $\theta$ considered random.
- Under the assumption iid, we have

  $[\theta|D]$=[D| $\theta$][$\theta$]/ $\int_{\Omega}$[D| $\theta$][$\theta$]d $\theta$ (Bayes)

  the denominator is often intractable.
- Example: rv x is sampled from

$[x|\mu] = \exp(-(x-\mu)^2)/\sqrt{2\pi}$

$[\mu|\mu_0, \sigma_0] \propto \exp(-(\mu-\mu_0)^2/2\sigma_0)$

$\theta = \mu$

$\mu_0$ and $\sigma_0$ are hyperparameters

# Bayes: Prior, likelihood, posterior

- Example: We want to calculate the conditional pdf [x | D] (prediction), where D is the data. We have:

  $[\mathrm{x} \mid \mathrm{D}] = \int_{\Omega} [\mathrm{x}, \mathrm{D}, \mu]/[\mathrm{D}]\mathrm{d}\mu = \int_{\Omega} [\mathrm{x} \mid \mu, D][\mu \mid \mathrm{D}]\, \mathrm{d}\mu.$

  If $[\mathrm{x} \mid \mu, D] = [\mathrm{x} \mid \mu] = Gaus(x; \mu, \sigma)$,

  $[\mu \mid \mathrm{D}] = [\mathrm{D} \mid \mu][\mu]/[\mathrm{D}]$,

  $[\mathrm{D} \mid \mu] = \prod_{\mathrm{i}=1}^{\mathrm{n}} Gaus(x_i; \mu, \sigma)$

  $[\mu] \sim Gaus(\mu; \mu_0, \sigma_0)$

  then we obtain

  $[\mathrm{x} \mid \mathrm{D}] = \int_{\Omega} Gaus(x; \mu, \sigma) \prod_{\mathrm{i}=1}^{\mathrm{n}} Gaus(x_i; \mu, \sigma)\ Gaus(\mu; \mu_0, \sigma_0)\mathrm{d}\mu/[D]$

# Hierarchical models

- Bayesian hierarchical model: Let the random quantities, Z (data), Y (hidden) and θ (parameter). The dimensions of these quantities can be high (e.g. 100 for Z, 1000 for Y, 5 for θ). We have :

  Data model $[Z \mid Y, \theta]$

  Process model $[Y \mid \theta]$

  Parameter model $[\theta]$

- The product of these three models gives $[Z \mid Y,\theta]\,[Y \mid \theta]\,[\theta]=[Z, Y, \theta]$
- To carry out the inference about Y and θ, we should use

  $$[Y, \theta \mid Z]=[Z \mid Y, \theta][Y \mid \theta][\theta]/[Z]$$

# Hierarchical models

- In the case of empirical Bayesian, $\theta$ is fixed but unknown (e.g. can be estimated by using the maximum likelihood).

  Data model $[Z \mid Y, \theta]$

  Process model $[Y \mid \theta]$

- In this case, the inference is made from $[Y \mid Z, \theta] = [Z \mid Y, \theta][Y \mid \theta] / [Z \mid \theta]$.
- The choice between an empirical Bayes model and a Bayes model depends on the data and the computational complexity. If few data are available, then $[\theta]$ should be considered random. However, the computational complexity is higher.

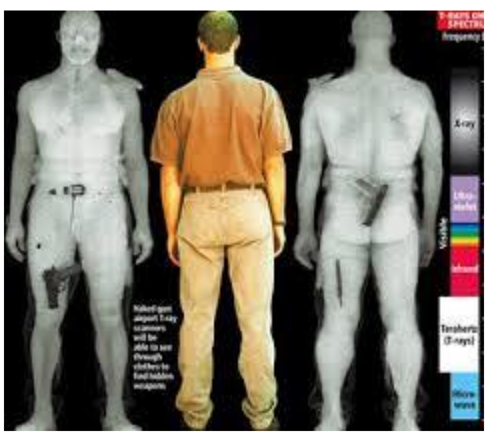
# Hierarchical models

- Example of terahertz image formation. This image may have 1024 bands.

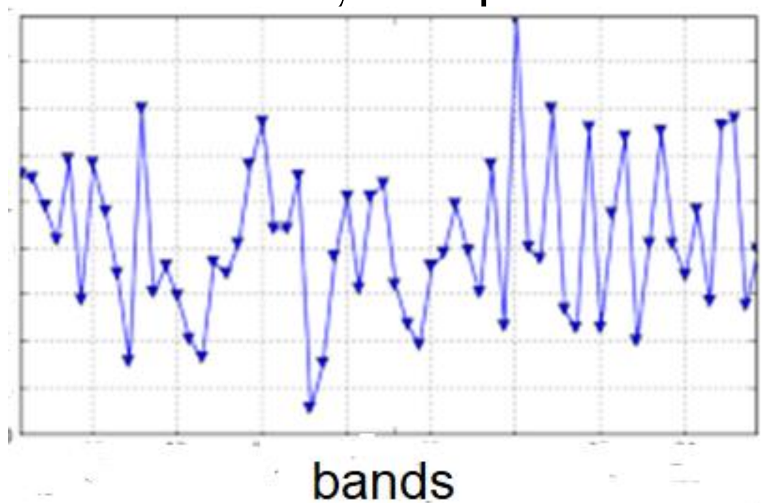
  Data model: $Z_n = H_n Y_n + \varepsilon_n$; $\varepsilon_n$ is Gauss$(0, a)$

  Process model: $Y_n = M Y_n + \eta_n$; $\eta_n$ is Gauss(0, b)

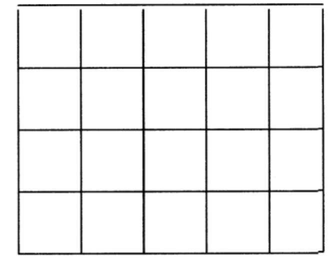
  Parameter model: $H_n$, M, a, and b.



Intensity at a pixel



# Hierarchical models



- Example: It is proposed to seek a submarine, that sank in the ocean (USS Scorpion 1968). The search area is divided into squares. Each square *i* is associated with a variable
  - $Z_i=1$ if the submarine was found in the square and 0 otherwise.
  - $Y_i=1$ if the submarine is in the square and 0 otherwise. $Y_i$ is not measurable (hidden).
- The probability of detection is $p_i=[Z_i=1 \mid Y_i=1]$
- The probability of occurrence $\pi_i=[Y_i=1]$
- Hierarchical Bayesian model
  - Data model $[Z_i \mid Y_i]=\mathrm{Ber}(Y_i p_i) = (Y_i p_i)^{z_i}(1-(Y_i p_i))^{1-z_i}$ and $[Z_i=0 \mid Y_i=0]=1$
  - Process model $[Y_i]=\mathrm{Ber}(\pi_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$

# Hierarchical models

- The probability that the submarine is in the square i and have not found is

$[Y_i=1 \mid Z_i=0]=[Z_i=0 \mid Y_i=1]\,[Y_i=1]/([Z_i=0 \mid Y_i=1]\,[Y_i=1]+[Z_i=0 \mid Y_i=0]\,[Y_i=0])$

Because $[Z_i=0 \mid Y_i=0]=1$ (no false detection)

$[Y_i=1 \mid Z_i=0]=(1-p_i)\,\pi_i/(1-p_i\,\pi_i)$

- The probability that the submarine is in the square *j* knowing he was not found in the square *i*

$[Y_j=1 \mid Z_i=0]=[Z_i=0 \mid Y_j=1]\,[Y_j=1]/([Z_i=0 \mid Y_i=1]\,[Y_i=1]+[Z_i=0 \mid Y_i=0]\,[Y_i=0])=\pi_j/(1-p_i\,\pi_i)$

The probabilities p and π are provided by experts.

# Model estimation/selection

- Let an iid data $D=\{x_1 \ldots x_n\}$ sampled from some pdf.
- We need to carry out the inference.
- Several choices
  - model estimation using maximum likelihood, EM…

    - $\max_\theta$ [D | $\theta$] $= \prod_{i=1}^{n} [x_i | \theta]$ (likelihood)

    - example (see normal law case studied before)
  - model selection using the MDL, AIC, MML…
  - the principle: $\max_\theta \ln [\theta \mid D] \propto \ln \prod_{i=1}^{n} [x_i | \theta] + \ln [\theta]$
  - example (AIC): $\min_\theta \ln [\theta \mid D] \approx -\ln \prod_{i=1}^{n} [x_i | \theta]$ + number of parameters to estimate
  - example (implementation in class)
  - model estimation using simulation.

# Model estimation/selection

- To estimate $[Y, \theta_p, \theta_D|Z]$ by simulation, we must run the following operations iteratively (Gibbs sampling – an MCMC method):

  simulate $[Y \mid \theta_p, \theta_D, Z]$

  simulate $[\theta_p \mid Y, \theta_D, Z]$

  simulate $[\theta_D \mid Y, \theta_p, Z]$

- An issue: Normalization constants (e.g. integral in the denominator) sometimes make it difficult the simulation.
- The solution (Gibbs-Metropolis): The constant (the integral) is not required for the simulation because we are using conditional probability. Example

  $[Y, \theta_p, \theta_D, Z] = f(Y, \theta_p, \theta_D, Z) = c\, g(Y, \theta_p, \theta_D, Z)$ and $c \neq 0$

  $f(Y \mid \theta_p, \theta_D, Z) = f(Y, \theta_p, \theta_D, Z) / \int f(Y, \theta_p, \theta_D, Z)dY = c\, g(Y, \theta_p, \theta_D, Z) / c\int g(Y, \theta_p, \theta_D, Z)dY = g(Y, \theta_p, \theta_D, Z) / \int g(Y, \theta_p, \theta_D, Z)dY = g(Y \mid \theta_p, \theta_D, Z)$

# Model estimation/selection

- Example of simulation, consider the first model:

$[Y \mid \theta_p, \theta_D, Z] = f(Y \mid \theta_p, \theta_D, Z) / \int f(Y \mid \theta_p, \theta_D, Z)\,dY$ where f is known, but the integral is not.

Let $y_c$ the value of Y computed previously and $y_s$ the simulated value by using f centered on $y_c$. There are two cases:

- If $f(y_s) > f(y_c)$ then we accept $y_s$ with the probability 1
- If $f(y_s) < f(y_c)$ then we accept $y_s$ with the probability $f(y_s)/f(y_c)$

In both cases, we accept $y_s$ with the probability $\min\left(1, \frac{f(y_s)}{f(y_c)}\right)$ and therefore we accept $y_c$ with the probability 1-$\min\left(1, \frac{f(y_s)}{f(y_c)}\right)$. The acceptance/rejection rule is

$$y_n = \begin{cases} y_s \; with\ probability\ \min(1, f(ys)/f(yc)) \\ y_c \; with\ probability\ 1 - \min(1, f(ys)/f(yc)) \end{cases}$$

# Model estimation/selection

Simulation algorithm

- Given data Z, $\theta_p^0$, $\theta_D^0$
- For n=1, *MaximumIteration* do
  - Generate a random number y from the pdf $[Y \mid \theta_p = \theta_p^{n-1}, \theta_D = \theta_D^{n-1}, Z]$
  - Rejection/acceptance decision rule to get $y^n$
  - Generate a random number $\theta_p$ from the pdf $[\theta_p \mid Y = y^n, \theta_D = \theta_D^{n-1}, Z]$
  - Rejection/acceptance decision rule to get $\theta_p^n$
  - Generate a random number $\theta_D$ from the pdf $[\theta_D \mid Y = y^n, \theta_p = \theta_p^n, Z]$
  - Rejection/acceptance decision rule to get $\theta_D^n$

End for

# Model estimation/selection

- The evaluation of the posterior probability can be done by:
  - the sampling of Z formY and $\theta$
  - the squared error between the actual data (Z) and synthetic data (previous step)

# Case study- attention allocation

- Attention allocation concerns the quantity of allocated cognitive resources to a subject, a theme, a location, an object… It is used in cognitive sciences, marketing…
- Can we understand the attention allocation to the body and emotion of young women with bulimia ?
- Experimental protocol
  - 20 pictures of unknown women with different weights and emotions.
  - 38 young women were tested. They are partitioned to two groups (18 suffering from bulimia and 20 not).
  - Every young woman labels each picture as X a lightweight and happy women or as Y a heavyweight and sad women. She provides a non negative score for weight and a non negative score for emotion. The total of the two scores must be equals to 10.

# Case study- attention allocation

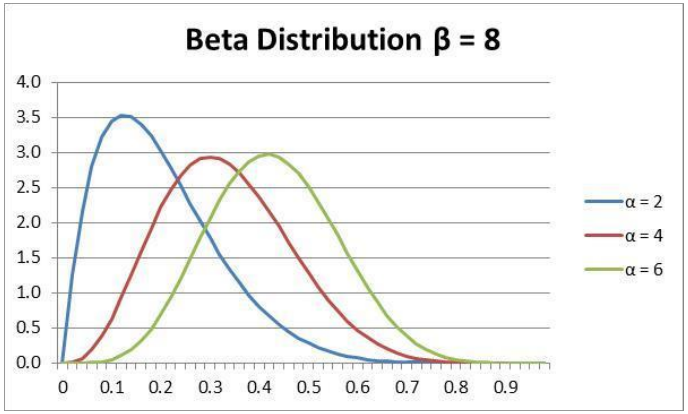
- Hierarchical model
  - we have N women and M images
  - $\mathbf{F}$=$F_1,\ldots,F_N$ , data vectors, where $F_n=(f_{n1}\ldots f_{nM})$ and $f_{nm}=x_{nm}$ the score for weight.
  - $W= W_1,\ldots,W_N$ , the relevance of the scores that the women gave.
  - $\boldsymbol{\theta} =\theta_1,\ldots, \theta_N$ , parameters
  - $\boldsymbol{\Psi}= \Psi_1, \Psi_2$, parameters
  - For the $n^{th}$ women
    - $[F_n \mid \theta_n, W_n]=\prod_{i=1}^{M} [f_{ni} \mid \theta_n, w_{ni}] = \prod_{i=1}^{M} g(f_{ni} \mid \mu_n, A_n, w_{ni})$
    - $[\theta_n \mid \theta^0]= [\mu_n \mid \theta_\mu^0] [A_n \mid \theta_A^0]$
  - For the N women $[F \mid \boldsymbol{\theta}, W, \boldsymbol{\theta}^0]=\prod_{i=1}^{N} [F_n \mid \theta_n, W_n] [\theta_n \mid \theta_n^0]$

# Case study- attention allocation

- The relevance is put into two classes

  $[W \mid \boldsymbol{\Psi}] = \prod_{n=1}^{N} (p(c=0)\ \mathrm{Beta}(w_n; \alpha_1, \beta_1) + p(c=1)\ \mathrm{Beta}(w_n; \alpha_2, \beta_2))$

  $[\boldsymbol{\Psi} \mid \Psi^0] = [\alpha, \beta \mid \alpha^0, \beta^0]$




- Inference
  - about $\boldsymbol{\theta}$, W, $\boldsymbol{\Psi}$
  - then the posterior is $[\boldsymbol{\theta}, W, \boldsymbol{\Psi} \mid \mathbf{F}, \boldsymbol{\theta}^0, \boldsymbol{\Psi}^0] = [\mathbf{F} \mid W, \boldsymbol{\theta}, \boldsymbol{\theta}^0]\ [\boldsymbol{\Psi} \mid \Psi^0]\ [\boldsymbol{\theta} \mid \boldsymbol{\theta}^0][W \mid \boldsymbol{\Psi}]$

# Case study- Image quality

- Given a collection of a color images and a set of users. Each user annotates (real number) the image according to its visual quality.


- Given an image, can we predict the subjective quality?

# Case study- Image quality

- Let x be a random variable where values are a feature of the image and r a subjective quality.

$[x,r\,|\,\Theta]=\Sigma_{k=0}^{M}\ \alpha_k\ [x,r\,|\,\theta_k]$

$[x,r\,|\,\theta_k]=[x\,|\,\varphi_k][r\,|\,x,\mu_k]$

$[\alpha\,|\,\ldots]$

$[\varphi_k\,|\,\ldots]$

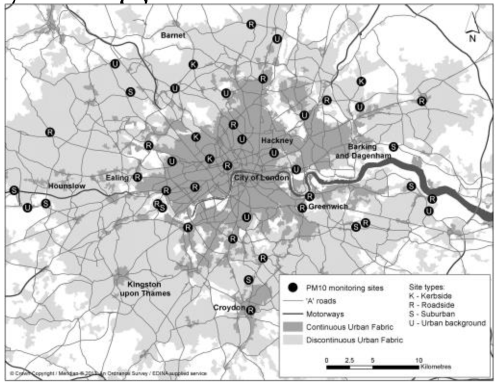
$[\mu_k\,|\,\ldots]$

- Inference

$[\Theta\,|\,r,x] \propto\ [x,r\,|\,\Theta][\Theta]$

# Case study- Short-term exposure to particle pollution in urban area

- Particles of interest are harmful for health. They originate from the transport of pollutants.




- Let y(s,t) the concentration of inhalable particle mater of diameter less than 10 microns.
- $[y(s,t)\,|\,\Theta] \sim \mathrm{Gauss}(y(s,t);\, \mu(s,t), \sigma(s))$
- $\mu(s,t)=a+b(t)$, where *a* is a residual mean concentration across the urban area and b(t) time-varying latent regional process.

# Case study- Short-term exposure to particle pollution in urban area

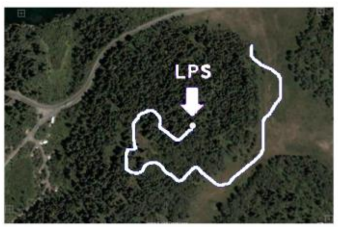
- Assuming that concentrations at the city scale derive largely from information borrowed from rural measurements.

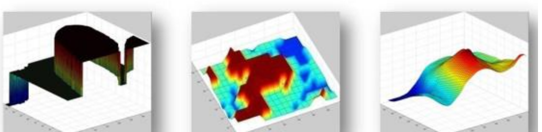
$[r(j,t)\,|\,\varphi] \sim \text{Gauss}(r(j,t);\, b(t), sd(j))$ is time series of pollution data from the rural site *j* measuring the long-range transport of particles proportion into the urban area.

- $[b(t)\,|\,\ldots]$
- $[\sigma(s)\,|\,\ldots]$
- $[sd(j)\,|\,\ldots]$
- Inference about $\varphi$ and $\Theta$

# Case study- A person in the forest

- We assume that in a forest, the path followed by an individual depends on the environment (topography, vegetation, and elevation), the person (experience, endurance, wayfinding), the context (goal, weather, visibility), and random phenomena (emotion…) .



Topology Vegetation Elevation

# Case study- A person in the forest

- [environment | Θ]
- [person, context | Ψ]
- [random phenomena | φ]
- [Θ | …]
- [Ψ | …]
- [φ | …]

# Case study- fairness in Machine learning

- Given the data

| | Male | Female |
|---|---|---|
| High income (C+) | 3256 | 590 |
| Low income (C-) | 7604 | 4831 |

If a bank uses these data for decision making about loans. Is the decision will be fair?

Probabilities related to women

[F]=(590+4831)/16281≈ 0.33

[F, C+]=590/16281 ≈ 0.04

[C+|F]=[C+,F]/[F] ≈ 0.11

[F|C+]=[C+,F]/[C+]=0.04/(0.2+0.04) ≈0.17

# Case study- fairness in Machine learning

| | Male | Female |
|---|---|---|
| High income (C+) | 3256 | 590 |
| Low income (C-) | 7604 | 4831 |

Probabilities related to men

$[M]=(3256+7604)/16281 \approx 0.67$

$[M,C+]=3256/16281 \approx 0.2$

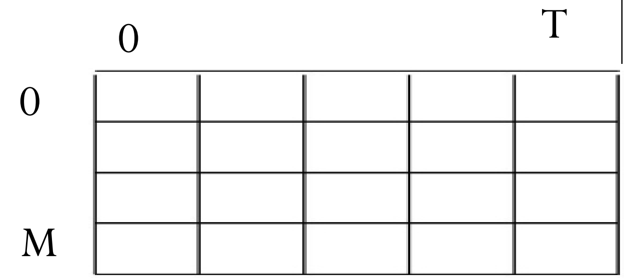
$[C+|M]=[C+,M]/[M] \approx 0.3$

$[M|C+]=[C+,M]/[C+]=0.2/(0.2+0.04) \approx 0.83$

- A discrimination measure $[C+|F]/[C+|M] \approx 0.37$
- Women will be disadvantaged for any decision based on these data.
- The genre and the incomes are correlated.

# Spatiotemporal reasoning

- Let
  - $D_s = \{0, 1, \ldots, M\}$ a spatial domain. A rv is associated with each position; $Y(0,t) \ldots Y(M,t)$.
  - $D_t = \{0, 1, \ldots, T\}$ temporal domain and the rv $Y(s_0,0) \ldots Y(s_0,T)$.




- A sequence of random variables is called random process.

# Spatiotemporal reasoning

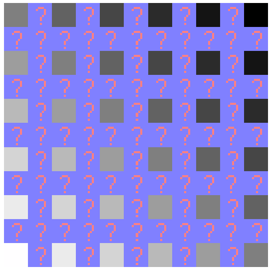
- Example of spatial process
  - it is proposed to find the missing gray levels in an image.

    $Y(i)=a_1 Y(i-1)+b_1 Y(i+1)+e_i$ , $i=1..M-1$

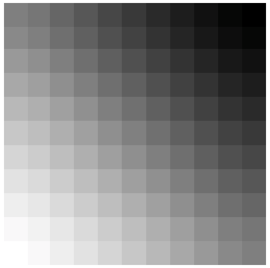
    $Y(0)=a_0\ e_0$

    $Y(M)=b_0\ e_{24}$

    $e_i \sim \text{Gauss}(0,\sigma_s)$



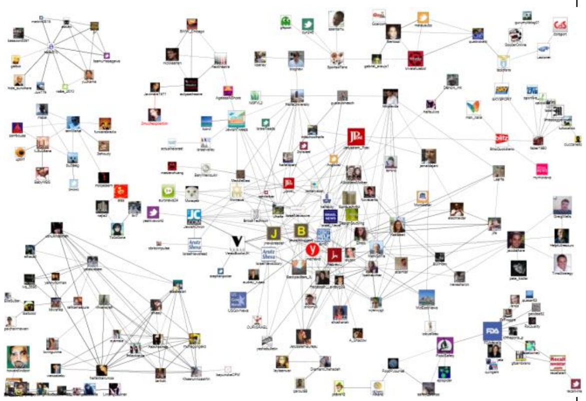
Before

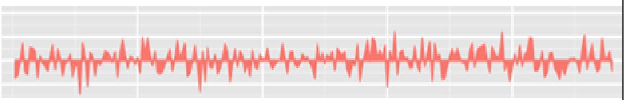


After

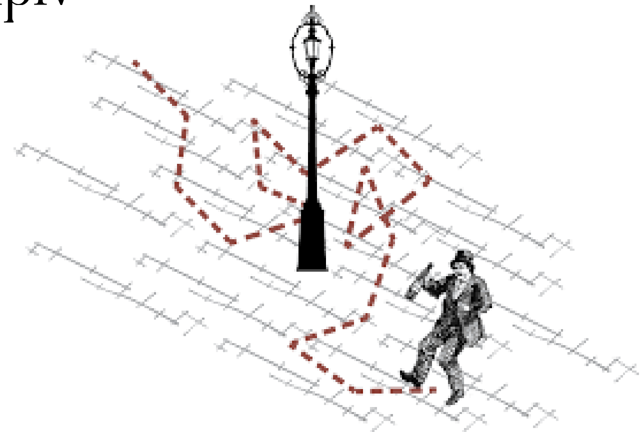
# Spatiotemporal reasoning

- Example of temporal process
  - we want to predict the centrality (sociability) of a user of a social network (graph); i.e. Y(t) is the number of incoming and outgoing arcs divided by the total number of edges in the graph.
  - $Y(t)=\alpha_1 Y(t\text{-}1)+\epsilon_t$, t=1..T et $Y(0)=\alpha_0\,\epsilon_0$; $\epsilon_t \sim \text{Gauss}(0, \sigma_t)$
  - Markov process of order 1 and the random walk are written this way. When $\alpha_1=1$, it simplv performs integration of noise $\epsilon_t$.





Centrality as function of time (time series)



# Spatiotemporal reasoning

- The dependence of random variables is different in the two previous processes.
- Example of spatiotemporal process
  - it is proposed to find the missing colors in a video

$Y(i,t)=a_1Y(i\text{-}1,t)+a_2Y(i+1,t)+a_3Y(i,t\text{-}1)+e_{it}$ , $i=1..M\text{-}1$, $t=1..T$

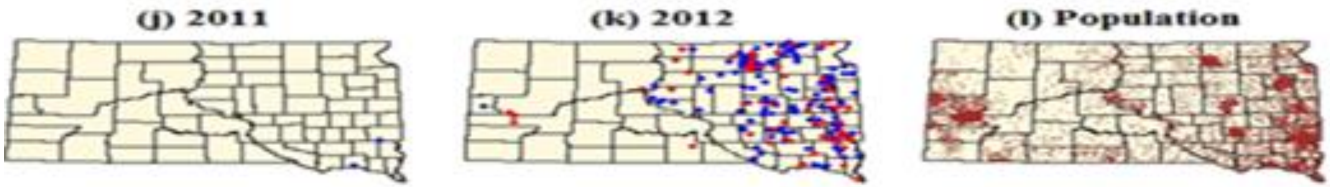
$Y(0,t)=a_0\ e_{0t}$

$Y(M,t)=b_0\ e_{Mt}$

$Y(i,0)=c_0\ e_{i0}$

$e_{it}\sim Gauss(0,\sigma)$

  - other uses …



# Bibliography


- N. Cressie and C.K. Wikle. Statistics for Spatio-Temporal Data. Wiley, 2011.
- R. O. Duda et al. Pattern Classification. Wiley-Interscience, 2001.
- S. Wang and P. Groth. Measuring the Dynamic Bi-directional Influence between Content and Social Networks. The 9th Int. Semantic Web Conference (ISWC2010).
- http://www.indiana.edu/~kruschke/articles/KruschkeVanpaemel2015.pdf
- T. Calders and S. Verwer. Three naive Bayes approaches for discrimination-free classification. Data Min Knowl Disc 21, pp. 277–292, 2010.
- T. Kamishima, S. Akaho, and J. Sakuma. Fairness-aware Learning through Regularization Approach. IEEE ICDMW, 2011.
- L. Lin and M. A. Goodrich. A Bayesian approach to modeling lost person behaviors based on terrain features in Wilderness Search and Rescue. Computational and Mathematical Organization Theory 16, pp 300–323, 2010.