Multi-componets Data, Signal and Image Processing for Biological and Medical Applications

Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes
UMR 8506 CNRS - CS - Univ Paris Sud
CentraleSupélec, Gif-sur-Yvette.

djafari@lss.supelec.fr
http://djafari.free.fr

January 6, 2017

Summary 1:

- Data, signals, images in Biological and medical applications
  - Individual cells, Population of cells, Small animals, Human
  - In vitro and In Vivo
- A great number of data, variables, time series, signals, images, ...
  - Genes expression, Hormones, temperature, ECG, EMG, ...
  - Tomographic images (X rays, PET, SPECT, IRM),
  - 3D body volume, fMRI, Holographic, multi- and Hyper-spectral images, ...
- Need for Visualization tools
  - multicomponent, multivariate and multidimensional
  - Time domain
  - Transformed domain: Fourier, Wavelets, Time-Frequency...
  - Scatter plots, histograms, statistics, ...

Summary 1.1: Data, Signal, Image

- Data, Signal, Image
  - Data: Unstructured
  - Signal: Structured in time
  - Image: Structured in space
  - Extensions: 3D, Space-Time, Space-Frequency, ...
- Different representations
  - Data: points in an abstract space, manifold
  - Signal: time and frequency
  - Image: space and spatial frequency
  - Extensions: 3D, Space-Time, Space-Frequency, ...

Summary 1.2: Data, unstructured

- One variable case
  - Histogram and Probability distribution
  - Parametric and Non parametric
  - Parametric models:
    - Method of moments
    - Maximum Likehood
    - Bayesian estimation
- Muti variable case
  - Joint Histogram and Joint Probability distribution
  - Correlation and Independence
  - Conditional and Marginal pdfs
  - Copula
  - Related estimation problems

Summary 2.1: Time series

- Time serie and Fourier representation
  - Continuous / Discrete
  - Correlation, Inter-correlation, Inter-dependance
  - Stationarity / non-stationarity
  - Convolution and Deconvolution
- Filtering and Denoising
- Modelling and Prediction
- Parametric and Non parametric models
- Parametric models:
  - Least Squares
  - Maximum Likehood
  - Bayesian estimation

Summary 2.2: Images

- Continuous / Discrete
- Gray and Color images
  - 2D FT and FFT
  - 2D Correlation and inter-correlation
  - Stationarity / non-stationarity
  - 2D Convolution
- Filtering and Denoising
- Modelling and Prediction
- Simple Markovian models
- Contours and Regions
- Hierarchical Markov models

Summary 3: Data redundancy, Dimensionality Reduction, ...

- Redundancy and structure
- Dimentionality Reduction
- PCA and ICA
- PPCA and its extensions
- Stationarity / non-stationarity
- Discriminant Analysis (DA)
- Classification and Clustering
- Mixture Models
- Factor Analysis
- Blind Sources Separation

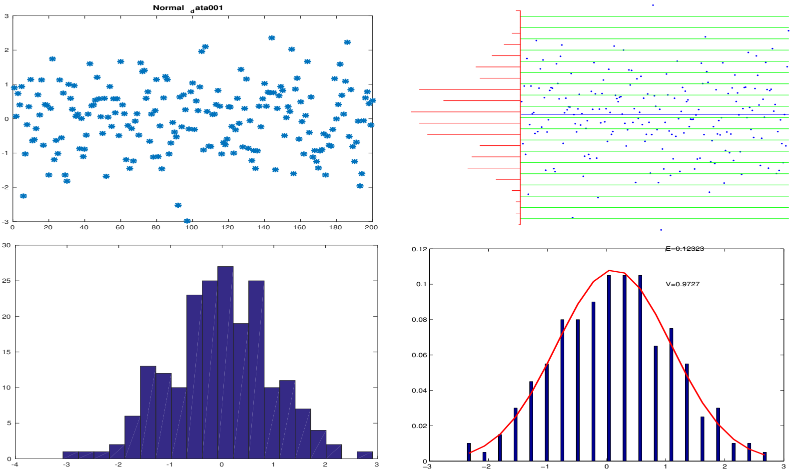
Summary 4 to 8: Medical and Biological Applications Case studies

- Signals:
  - Electro-Cardio-Gram (ECG)
  - Electro-Encephalo-Gram (EEG)
  - Electro-Myo-Gram (EMG)
  - Magneto-Electro-Gram (MEG)
- Images:
  - Computed Tomography: X ray CT Scan
  - Magnetic Resonance Imaging (MRI)
  - Positon Emission Tomography (PET)
  - Single Photon Emission Computed Tomography (SPECT)
  - Phosphorescence, Molecular Imaging
- Case studies in Cancer Research

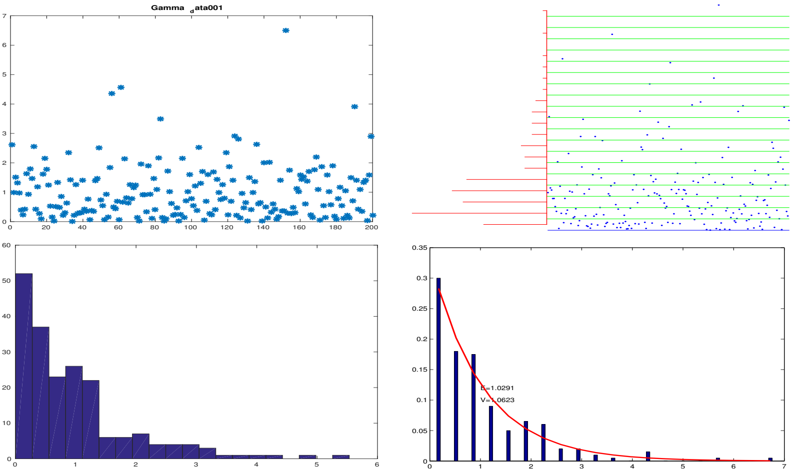
1D data: one variable

- Data: $x_i, i = 1, \cdots, M$
- 1D plot, mean, median, variance
- No order: exchangeable
- histogram, probability distribution,
- Statistical modelling: expected value, variance, mode, median, Higher order moments, entropy
- Parametric, semi-parametric and Non Parametric modelling
- Parameter estimation: MM, ML, Bayesian
- Model selection: AIC, BIC, ...

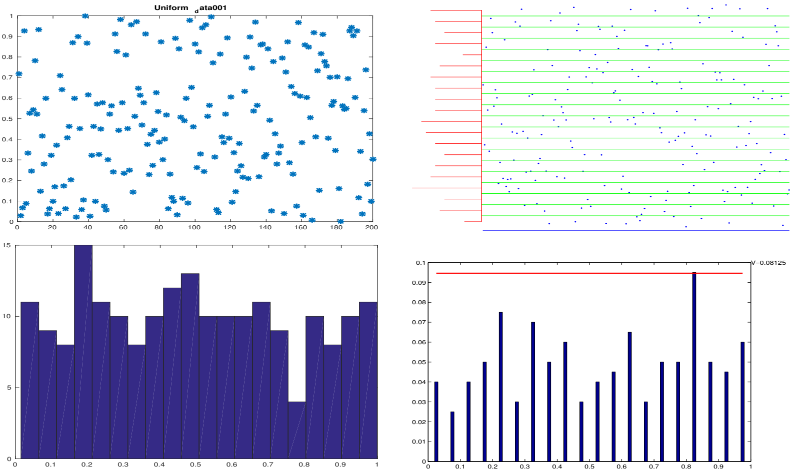
1D data (Gaussian)



1D data (Gamma)



1D data (Uniform)

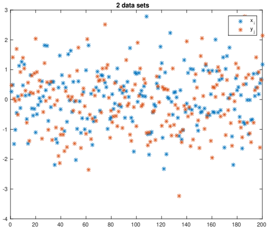


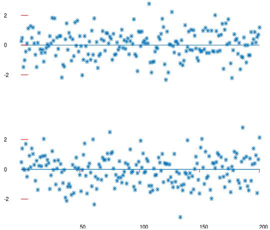
1D data: Statistics and probability modeling

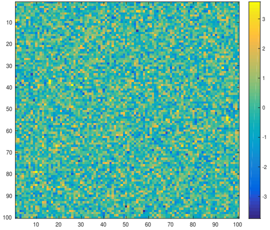
- Statistics
  - Mean, Variance, standard deviation, moments,...
- Histogram and probability distribution matching
  - Uniform, Gaussian, Gamma,... shapes
- Probabilistic modelling
  - Method of Moments (MM)
  - Maximum Entropy (ME)
  - Maximum Likelihood (ML)
  - Bayesian estimation
  - Non Parametric Bayesian methods

Multi-component, multi-variate, multi-dimensional data

- bi-component: $\{(x_i, y_i)\}, i = 1, \cdots, M\}$, $2M$ elements
- bi-variate: $\{x_i, i = 1, \cdots, M\}, \{y_j, j = 1, \cdots, N\}$, $M + N$ elements
- bi-dimensional: Images: $x_{i,j}, i = 1, \cdots, M, j = 1, \cdots, N$, $M * N$ elements

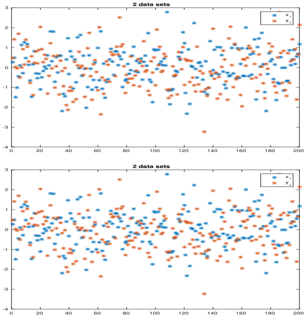


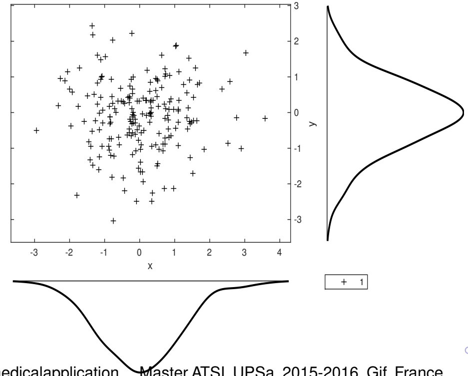




Bi-component or Bi-variate data

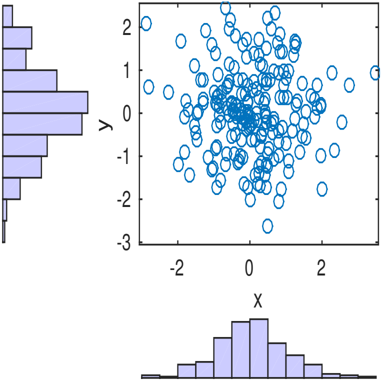
- 2D distribution: joint probability distribution $p(x, y)$
- Conditionals $p(x|y), p(y|x)$
- Marginal distributions $p(x)$, $p(y)$
- Expected values $E(X), E(Y)$, variances $V(X), V(Y)$, and Covariances, Higher order moments, entropy
- Independence tests
- Copula, ...

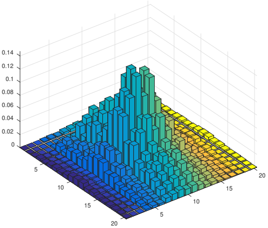


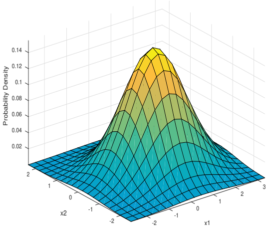


Bivariate data

- Joint, marginals and conditional probability density functions







Probability theory Review: Discrete and Continuous variables probability laws

- What is a probability?
- What is a random variable?
- What are the main rules of probability
- Discrete probability laws:
  - Bernouilli
  - Binomial
  - Poisson

Continuous probability laws:

- Uniform $U(.|a,b)$
- Beta $\mathcal{B}(.|\alpha,\beta)$
- Gaussian $\mathcal{N}(.|\mu,v)$
- Generalized Gaussian $\mathcal{GG}(.|\gamma,\beta)$
- Gamma $\mathcal{G}(.|\alpha,\beta)$
- Student-t $\mathcal{S}(.|\nu,\mu,\lambda)$
- Cauchy $\mathcal{C}(.|\mu,\lambda)$

Uniform and Beta distributions

- Uniform:

$$X \sim \mathrm{U}(.|a,b) \longrightarrow p(x) = \frac{1}{b-a}, \quad x \in [a,b]$$

$$\mathrm{E}\{X\} = \frac{a+b}{2}, \quad \mathrm{Var}\{X\} = \frac{(b-a)^2}{12}$$

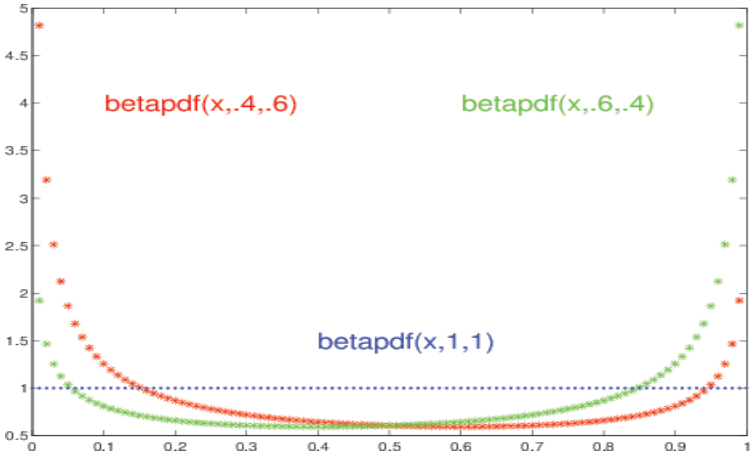
- Beta:

$$X \sim \mathrm{Beta}(.|\alpha,\beta) \longrightarrow p(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, x \in [0,1]$$

$$\mathrm{E}\{X\} = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{Var}\{X\} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

- $\mathrm{Beta}(.|1,1) = \mathrm{U}(.|0,1)$

Uniform and Beta distributions



Gaussian distributions

Different notations:

- classical one with mean and variance:

$$X \sim \mathcal{N}(.|\mu, \sigma^2) \longrightarrow p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

$$\mathrm{E}\{X\} = \mu, \quad \mathrm{Var}\{X\} = \sigma^2$$

- mean and precision parameters:

$$X \sim \mathcal{N}(.|\mu, \lambda) \longrightarrow p(x) = \frac{\lambda}{\sqrt{2\pi}} \exp\left[-\frac{\lambda}{2}(x-\mu)^2\right]$$

$$\mathrm{E}\{X\} = \mu, \quad \mathrm{Var}\{X\} = \sigma^2 = \frac{1}{\lambda}$$

Generalized Gaussian distributions

- Gaussian:

$$X \sim \mathcal{N}(.|\mu,\sigma^2) \longrightarrow p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2\right]$$

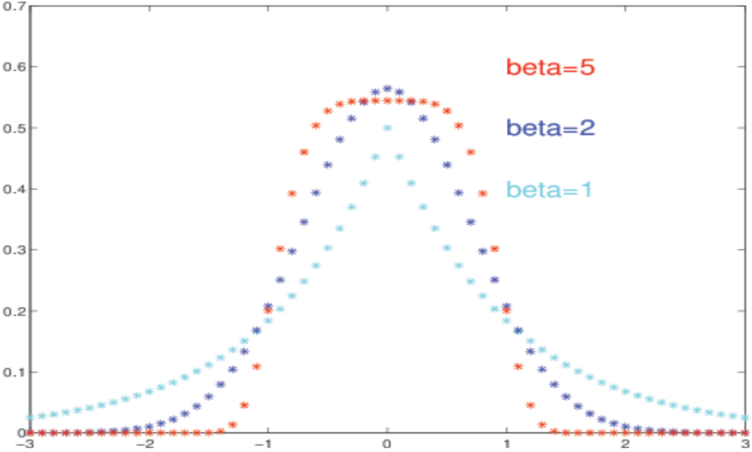
- Generalized Gaussian:

$$X \sim \mathcal{GG}(.|\alpha,\beta) \longrightarrow p(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left[-\left(\frac{|x-\mu|}{\alpha}\right)^\beta\right]$$

$$\mathrm{E}\{X\} = \mu, \quad \mathrm{Var}\{X\} = \frac{\alpha^2\Gamma(3/\beta)}{\gamma(1/\beta)}$$

- $\beta > 0$, $\beta = 1$ Laplace, $\beta = 2$: Gaussian, $\beta \mapsto \infty$: Uniform

Gaussian and Generalized Gaussian distributions



Gamma distributions

- Forme 1:

$$p(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \text{ for } x \geq 0$$

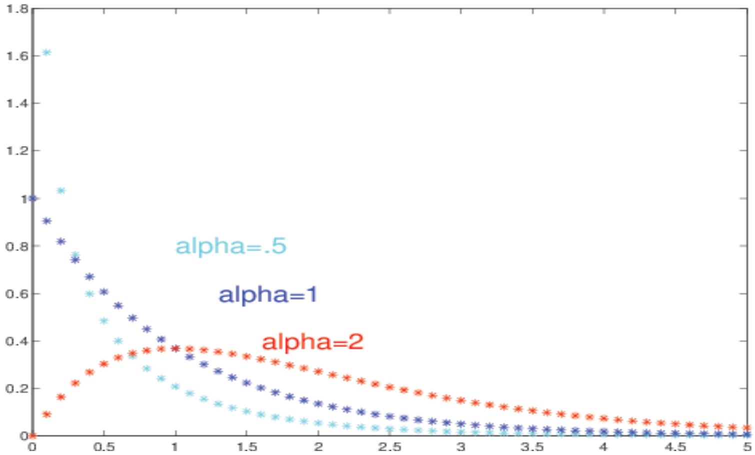
$$\mathrm{E}\left\{X\right\} = \frac{\alpha}{\beta}, \quad \mathrm{Var}\left\{X\right\} = \frac{\alpha}{\beta^2}, \quad \mathrm{Mod}(X) = \frac{\alpha-1}{\alpha+\beta-2}$$

- Forme 2: $\theta = 1/\beta$

$$p(x|\alpha,\theta) = \frac{\theta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \text{ for } x \geq 0$$

- $\alpha = 1$: Exponential,
- $0 < \alpha < 1$ decreasing,
- $\alpha > 1$ Mode=$\frac{\alpha-1}{\beta}$

Gamma distributions



Student-t and Cauchy distributions

- Student's t-distribution has the probability density function:

$$p(x|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} = \frac{1}{\sqrt{\nu}\,B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

  where
  - ν is the number of degrees of freedom,
  - Γ is the Gamma function and
  - B is the Beta function.
- $\nu = 1$ gives Cauchy distribution.

$$p(x) = \frac{\pi}{1+x^2}$$

- Cauchy distribution:

$$p(t|\mu) = \frac{\pi}{1+(x-\mu)^2}$$

Student-t and Cauchy distributions

- Three parameters location (μ) / scale (λ) / degree of freedom (ν) version

$$p(x|\mu,\lambda,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\frac{\lambda}{\pi\nu}\right)^{\frac{1}{2}} \left[1 + \frac{\lambda(x-\mu)^2}{\nu}\right]^{-\frac{\nu+1}{2}}$$

$$\begin{array}{lll} \mathrm{E}\{X\} & = \mu & \text{for } \nu > 1, \\ \mathrm{Var}\{X\} & = \frac{1}{\lambda}\frac{\nu}{\nu-2} & \text{for } \nu > 2, \\ \mathrm{mode}(X) & = \mu. & \end{array}$$

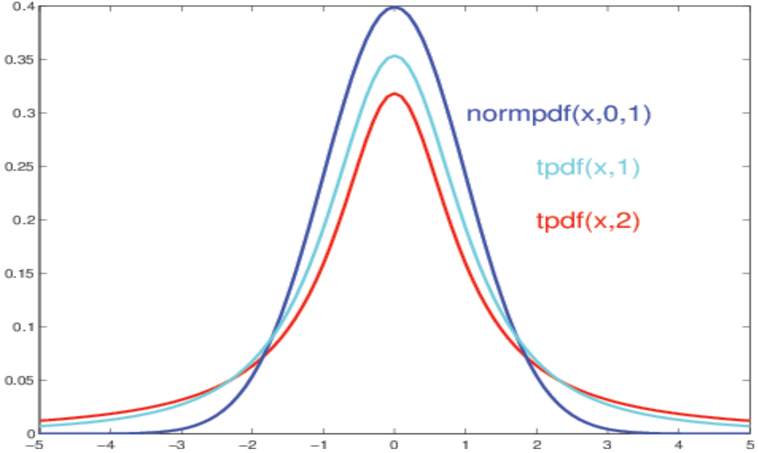
- Interesting relation between Student-t, Normal and Gamma distributions:

$$\mathcal{S}(x|\mu,1,\nu) = \int \mathcal{N}(x|\mu,1/\lambda)\,\mathcal{G}(\lambda|\nu/2,\nu/2)\,\mathrm{d}\lambda$$

$$\mathcal{S}(x|0,1,\nu) = \int \mathcal{N}(x|0,1/\lambda)\,\mathcal{G}(\lambda|\nu/2,\nu/2)\,\mathrm{d}\lambda$$

Student and Cauchy

$$p(x|\nu) \propto \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$



Multivariate continuous probability laws

- Gaussian $\mathcal{N}(.|\boldsymbol{\mu}, \boldsymbol{V})$
- Student-t $\mathcal{S}(.|\gamma, \beta)$
- Hyperbolic $\mathcal{H}(.|\gamma, \beta)$

Multivariate Gaussian

Different notations:

- mean and covariance matrix (classical): $\boldsymbol{X} \sim \mathcal{N}(.|\boldsymbol{\mu}, \boldsymbol{\sigma})$

$$p(\boldsymbol{x}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

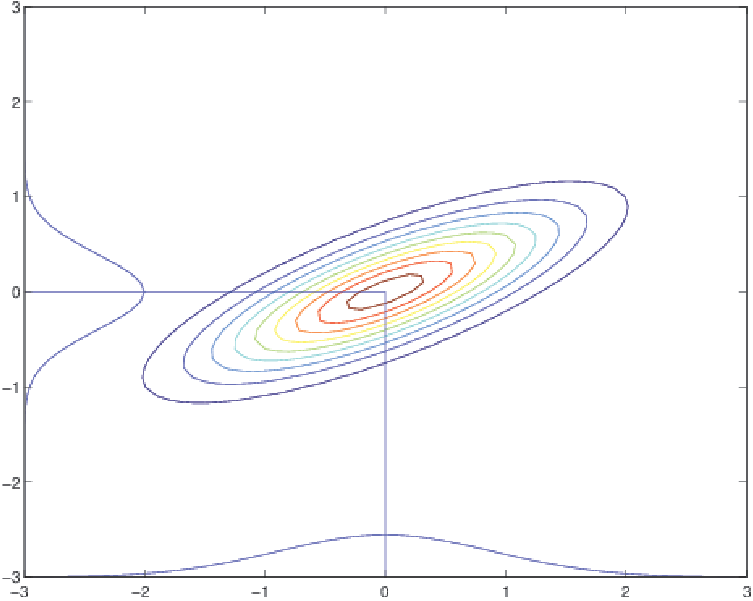
$$\mathrm{E}\{\boldsymbol{X}\} = \boldsymbol{\mu}, \quad \mathrm{cov}[\boldsymbol{X}] = \boldsymbol{\Sigma}$$

- mean and precision matrix: $\boldsymbol{X} \sim \mathcal{N}(.|\boldsymbol{\mu}, \boldsymbol{\Lambda})$

$$p(\boldsymbol{x}) = (2\pi)^{-n/2}|\boldsymbol{\Lambda}|^{1/2} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Lambda}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

$$\mathrm{E}\{\boldsymbol{X}\} = \boldsymbol{\mu}, \quad \mathrm{cov}[\boldsymbol{X}] = \boldsymbol{\Lambda}^{-1}$$

Multivariate normal distributions



Multivariate Student-t

$$p(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) \propto |\boldsymbol{\Sigma}|^{-1/2}\left[1+\frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{(\nu+p)/2}$$

- $p=1$

$$f(t)=\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}}(1+t^2/\nu)^{\frac{-(\nu+1)}{2}}$$

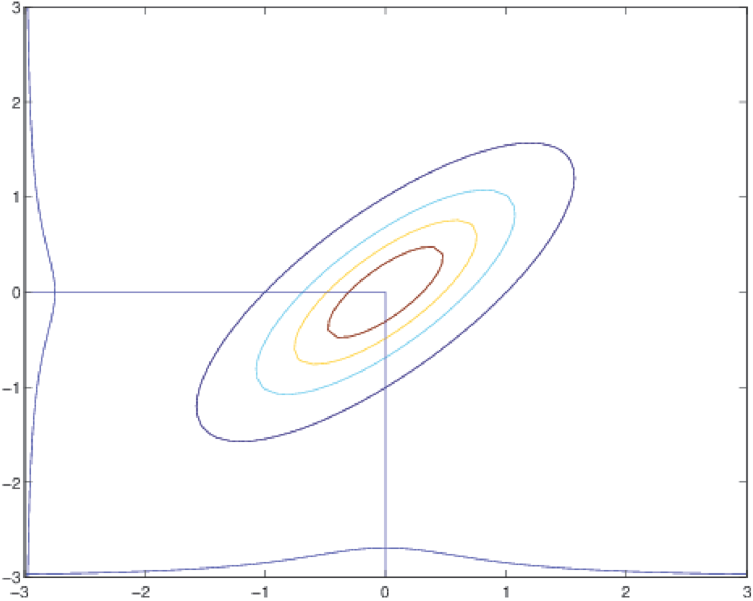
- $p=2$, $\boldsymbol{\Sigma}^{-1}=\boldsymbol{A}$

$$f(t_1,t_2)=\frac{\Gamma((\nu+p)/2)}{\Gamma(\nu/2)\sqrt{\nu^p\pi^p}}\frac{|\boldsymbol{A}|^{1/2}}{2\pi}\left(1+\sum_{i=1}^{p}\sum_{j=1}^{p}A_{ij}t_i\,t_j/\nu\right)^{\frac{-(\nu+2)}{2}}$$

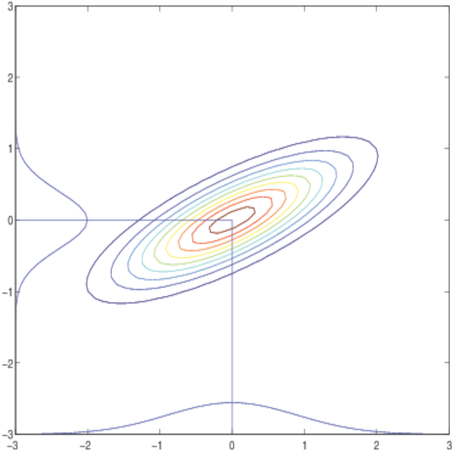
- $p=2$, $\boldsymbol{\Sigma}=\boldsymbol{A}=\boldsymbol{I}$

$$f(t_1,t_2)=\frac{1}{2\pi}(1+(t_1^2+t_1^2)/\nu)^{\frac{-(\nu+2)}{2}}$$

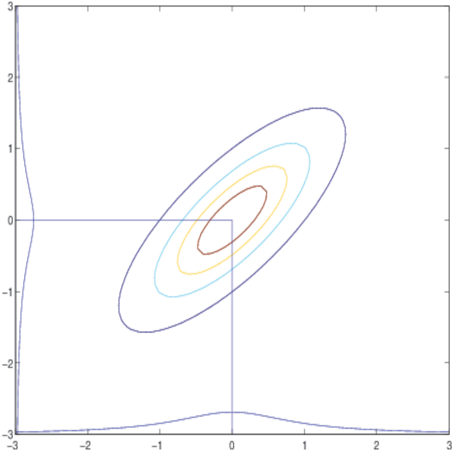
Multivariate Student-t distributions



Multivariate normal distributions



Normal



Student-t

Multivariate elliptic

$$p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \propto g\left(1 + \frac{1}{\nu}(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right)$$

- $g(z) = \exp\left[-z/2\right] \longrightarrow$ Multivariate normal
- More general: Caracteristic function

$$\exp\left[i\boldsymbol{x}'\boldsymbol{\mu}\right]\Psi(\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x})$$

Multivariate Wishart

Suppose $\boldsymbol{X}$ is an $n \times p$ matrix, each row of which is independently drawn from a p-variate normal distribution with zero mean:

$$\boldsymbol{X}_i = [x_i^1, \ldots, x_i^p] \sim \mathcal{N}_p(.|0, \boldsymbol{V})$$

Then the Wishart distribution is the probability distribution of the $p \times p$ random matrix $\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{X}$ known as the scatter matrix:

$$\boldsymbol{S} \sim \mathcal{W}_p(.|\boldsymbol{V}, \nu).$$

- The positive integer ν is the number of degrees of freedom.
- Sometimes this is written $\mathcal{W}(\boldsymbol{V}, p, \nu)$.
- For $n \geq p$ the matrix $\boldsymbol{S}$ is invertible with probability 1 if $\boldsymbol{V}$ is invertible.
- If $p = 1$ and $V = 1$ then this distribution is a chi-squared distribution with ν degrees of freedom

Multivariate Wishart

- probability density:

$$p(\boldsymbol{S}|\boldsymbol{V},\nu) = \frac{|\boldsymbol{S}|^{(\nu-p-1)/2} \exp\left[-\frac{1}{2}\mathrm{Tr}(\boldsymbol{V}^{-1}\boldsymbol{S})\right]}{2^{\nu p/2}\,|\boldsymbol{V}|^{\nu/2}\,\Gamma_p(\nu/2)}$$

where $\Gamma_p(.)$ is the multivariate gamma function defined as

$$\Gamma_p(\nu/2) = \pi^{p(p-1)/4}\Pi_{j=1}^{p}\Gamma\left[\nu/2+(1-j)/2\right].$$

$$\mathrm{E}\left\{\boldsymbol{S}\right\} = \boldsymbol{V}, \quad \mathrm{Var}\left\{\boldsymbol{S}_{ij}\right\} = \nu(v_{ij}^2 + v_{ii}v_{jj})$$

Parameter estimation

We observe n samples $\boldsymbol{x} = \{x_1, \cdots, x_n\}$ of a quantity X whose pdf depends on certain parameters $\boldsymbol{\theta}$: $p(x|\boldsymbol{\theta})$.
The question is to determine $\boldsymbol{\theta}$.

- Moments method:

$$\mathrm{E}\left\{x^k\right\} = \int x^k p(x|\boldsymbol{\theta}) \, \mathrm{d}x \approx \frac{1}{n} \sum_{i=1}^{n} x_i^k, \quad k = 1, \cdots, K$$

- Maximum Likelihood

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} p(x_i|\boldsymbol{\theta}) \text{ or } \ln \mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln p(x_i|\boldsymbol{\theta})$$

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{\mathcal{L}(\boldsymbol{\theta})\} = \arg\min_{\boldsymbol{\theta}} \{-\ln \mathcal{L}(\boldsymbol{\theta})\}$$

- Bayesian approach

Bayesian Parameter estimation

- Likelihood

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{n} p(x_i|\boldsymbol{\theta})$$

- A priori

$$p(\boldsymbol{\theta})$$

- A posteriori

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

- Infer on $\boldsymbol{\theta}$ using $p(\boldsymbol{\theta}|\boldsymbol{x})$.
  For example:
  - Maximum A Posteriori (MAP)

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\boldsymbol{x})\}$$

  - Posterior Mean

$$\widehat{\boldsymbol{\theta}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{x}) \, \mathrm{d}\boldsymbol{\theta}$$

Parameter estimation: Normal distribution

$$p(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$p(\mu,\sigma|\boldsymbol{x}) = \frac{p(\mu,\sigma)}{p(\boldsymbol{x})} \prod_{i=1}^{N} p(x_i|\mu,\sigma)$$

$$p(\mu,\sigma|\boldsymbol{x}) = \frac{p(\mu,\sigma)}{p(\boldsymbol{x})} \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\sum_{i=1}^{N} \frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and} \quad s^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i-\bar{x})^2$$

$$p(\mu,\sigma|\boldsymbol{x}) = \frac{p(\mu,\sigma)}{p(\boldsymbol{x})} \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{(\mu-\bar{x})^2+s^2}{2\sigma^2/N}\right]$$

Parameter estimation: Normal distribution: σ known

- σ known: $p(\mu,\sigma) = p(\mu)\,\delta(\sigma-\sigma_0)$

$$
\begin{aligned}
p(\mu|\boldsymbol{x}) &= \frac{p(\mu)}{p(\boldsymbol{x})}\frac{1}{\left(2\pi\sigma_0^2\right)^{N/2}} \exp\left[-\sum_{i=1}^N \frac{(x_i-\mu)^2}{2\sigma_0^2}\right] \\
&= \frac{p(\mu)}{p(\boldsymbol{x})}\frac{1}{\left(2\pi\sigma_0^2\right)^{N/2}} \exp\left[-\frac{(\mu-\bar{x})^2+s^2}{2\sigma_0^2/N}\right] \\
&\propto p(\mu) \exp\left[-\frac{(\mu-\bar{x})^2}{2\sigma_0^2/N}\right]
\end{aligned}
$$

- $p(\mu) = c \longrightarrow p(\mu|\boldsymbol{x}) = \mathcal{N}(\bar{x}, \sigma_0^2/N)$

$$
\mu = \bar{x} \pm \frac{\sigma_0}{\sqrt{N}}
$$

- $p(\mu) = \mathcal{N}(\mu_0, v_0) \longrightarrow p(\mu|\boldsymbol{x}) = \mathcal{N}(\widehat{\mu}, \hat{v})$

$$
\widehat{\mu} = \frac{v_0}{v_0+\sigma_0^2}\bar{x} + \frac{\sigma_0^2}{v_0+\sigma_0^2}\mu_0, \qquad \widehat{v} = \frac{v_0+\sigma_0^2}{v_0\sigma_0^2}
$$

Parameter estimation: Normal distribution

- σ not known:
- A first choice for prior

$$p(\mu, \sigma) \propto \begin{cases} C & \text{for} \quad \sigma > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- Another popular choice: uniform in μ and in $\log \sigma$.

$$p(\mu|\boldsymbol{x}) = \int_0^\infty \mathrm{d}\sigma \, p(\mu, \sigma|\boldsymbol{x}) \propto \int_0^\infty \mathrm{d}\sigma \, \frac{1}{\sigma^N} \exp\left[-\frac{(\mu - \bar{x})^2 + s^2}{2\sigma^2/N}\right]$$

Change variables to $t = 1/\sigma$, then

$$p(\mu|\boldsymbol{x}) \propto \int_0^\infty dt \, t^{N-2} \exp\left[-\frac{t^2}{2} N \left((\mu - \bar{x})^2 + s^2\right)\right] .$$

Repeated integrations by parts lead to

$$p(\mu|\boldsymbol{x}) \propto \left[N \left((\mu - \bar{x})^2 + s^2\right)\right]^{-\frac{N-1}{2}} ,$$

Student-t distribution.

$$\langle \mu \rangle = \bar{x} .$$

Conjugate priors

- The conjugate prior concept is tightly related to the sufficient statistics and exponential families.
- When $X \sim P_\theta(x)$, a function $h(X)$ is said to be a sufficient statistics for $\{P_\theta(x), \theta \in \mathcal{T}\}$ if the distribution of X conditioned on $h(X)$ does not depend on θ for $\theta \in \mathcal{T}$.
- A function $h(X)$ is said to be minimal sufficient for $\{P_\theta(x), \theta \in \mathcal{T}\}$ if it is a function of every other sufficient statistics for $P_\theta(x)$.
- A minimal sufficient statistics contains the whole information brought by the observation $X = x$ about θ.
- Suppose that $\{P_\theta(x), \theta \in \mathcal{T}\}$ has a corresponding family of densities $\{p_\theta(x), \theta \in \mathcal{T}\}$. A statistic T is sufficient for θ if and only if there exist functions g_θ and h such that

$$p_\theta(x) = g_\theta(T(x))\, h(x)$$

for all $x \in \Gamma$ and $\theta \in \mathcal{T}$.

Conjugate priors examples

- If $X \sim \mathcal{N}(\theta, 1)$ then $T(x) = x$ can be chosen as a sufficient statistics.
- If $\{X_1, X_2, \ldots, X_n\}$ are i.i.d. and $X_i \sim \mathcal{N}(\theta, 1)$ then

$$f(\boldsymbol{x}|\theta) = (2\pi)^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right]$$

$$= \exp\left[-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right] (2\pi)^{-n/2} \exp\left[-\frac{n}{2}\theta^2\right] \exp\left[\theta \sum_{i=1}^{n} x_i\right]$$

and we have $T(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$.

- Note that, in this case, we need to know n and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Note also that we can write

$$f(\boldsymbol{x}|\theta) = a(\boldsymbol{x})\, g(\theta) \exp\left[\theta\, T(\boldsymbol{x})\right]$$

where

$$g(\theta) = (2\pi)^{-n/2} \exp\left[-\frac{n}{2}\theta^2\right] \quad \text{and} \quad a(\boldsymbol{x}) = \exp\left[-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right]$$

Conjugate priors examples

- If $X \sim \mathcal{N}(0, \theta)$ then $T(x) = x^2$ can be chosen as a sufficient statistics.
- If $\{X_1, X_2, \ldots, X_n\}$ are i.i.d. and $X_i \sim \mathcal{N}(\theta_1, \theta_2)$ then

$$f(\boldsymbol{x}|\theta_1, \theta_2) = (2\pi)^{-n/2}\, \theta_2^{-1/2} \exp\left[-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)^2\right]$$

$$= (2\pi)^{-n/2}\, \theta_2^{-1/2} \exp\left[-\frac{n\theta_1^2}{2\theta_2}\right] \exp\left[-\frac{1}{2\theta_2}\sum_{i=1}^{n} x_i^2 + \frac{\theta_1}{\theta_2}\sum_{i=1}^{n} x_i\right]$$

and we have $T_1(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$ and $T_2(\boldsymbol{x}) = \sum_{i=1}^{n} x_i^2$.

Conjugate priors examples

- If $\{X_1, X_2, \ldots, X_n\}$ are i.i.d. and $X_i \sim \mathcal{N}(\theta_1, \theta_2)$ then

$$f(\boldsymbol{x}|\theta_1, \theta_2) = (2\pi)^{-n/2}\, \theta_2^{-1/2} \,\exp\left[-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)^2\right]$$

$$= (2\pi)^{-n/2}\, \theta_2^{-1/2} \,\exp\left[-\frac{n\theta_1^2}{2\theta_2}\right] \,\exp\left[-\frac{1}{2\theta_2}\sum_{i=1}^{n} x_i^2 + \frac{\theta_1}{\theta_2}\sum_{i=1}^{n} x_i\right]$$

and we have $T_1(\boldsymbol{x}) = \sum_{i=1}^n x_i$ and $T_2(\boldsymbol{x}) = \sum_{i=1}^n x_i^2$.

$$f(\boldsymbol{x}|\theta) = a(\boldsymbol{x})\, g(\theta_1, \theta_2) \exp\left[\frac{\theta_1}{\theta_2} T_1(\boldsymbol{x}) - \frac{1}{2\theta_2} T_2(\boldsymbol{x})\right]$$

$$g(\theta_1, \theta_2) = (2\pi)^{-n/2}\theta_2^{-1/2} \exp\left[-\frac{n\theta_1^2}{2\theta_2}\right] \quad \text{and} \quad a(\boldsymbol{x}) = 1.$$

- $\frac{\theta_1}{\theta_2}$ and $\frac{-1}{2\theta_2}$ are called canonical parametrization.
- It is also usual to use n: $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ and $\overline{x^2} = \frac{1}{n}\sum_{i=1}^n x_i^2$ as the sufficient statistics.

Conjugate priors examples

- If $X \sim \mathcal{G}(\alpha, \theta)$ then $T(x) = x$ can be chosen as a sufficient statistics.
- If $X \sim \mathcal{G}(\theta, \beta)$ then $T(x) = \ln x$ can be chosen as a sufficient statistics.
- If $X \sim \mathcal{G}(\theta_1, \theta_2)$ then $T_1(x) = \ln x$ and $T_2(x) = x$ can be chosen as a set of sufficient statistics.
- If $\{X_1, X_2, \ldots, X_n\}$ are i.i.d. and $X_i \sim \mathcal{G}(\theta_1, \theta_2)$ then it is easy to show that $T_1(\boldsymbol{x}) = \sum_{i=1}^n \ln x_i$ and $T_2(\boldsymbol{x}) = \sum_{i=1}^n x_i$.

Exponential families and Conjugate priors

- A class of distributions $\{P_\theta(\boldsymbol{x}), \theta \in \mathcal{T}\}$ is said to be an exponential family if there exist: $a(\boldsymbol{x})$ a function of Γ on R, $g(\boldsymbol{\theta})$ a function of $\mathcal{T}$ on R^+, $\phi_k(\boldsymbol{\theta})$ functions of $\mathcal{T}$ on R, and $h_k(\boldsymbol{x})$ functions of Γ on R such that

$$
\begin{aligned}
p_\theta(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{\theta}) &= a(\boldsymbol{x})\, g(\boldsymbol{\theta}) \exp\left[\sum_{k=1}^{K} \phi_k(\boldsymbol{\theta})\, h_k(\boldsymbol{x})\right] \\
&= a(\boldsymbol{x})\, g(\boldsymbol{\theta}) \exp\left[\boldsymbol{\phi}^t(\boldsymbol{\theta}) \boldsymbol{h}(\boldsymbol{x})\right]
\end{aligned}
$$

  for all $\theta \in \mathcal{T}$ and $x \in \Gamma$.

- This family is entirely determined by $a(\boldsymbol{x})$, $g(\boldsymbol{\theta})$, and $\{\phi_k(\boldsymbol{\theta}), h_k(\boldsymbol{x}),\ k = 1, \cdots, K\}$ and is noted $\mathbf{Exfn}(\boldsymbol{x}|a, g, \boldsymbol{\phi}, \boldsymbol{h})$

Exponential families and Conjugate priors

- When $a(\boldsymbol{x}) = 1$ and $g(\boldsymbol{\theta}) = \exp\left[-b(\boldsymbol{\theta})\right]$ we have

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\left[\boldsymbol{\phi}^t(\boldsymbol{\theta})\boldsymbol{h}(\boldsymbol{x}) - b(\boldsymbol{\theta})\right]$$

- Natural exponential family:
When $a(\boldsymbol{x}) = 1$, $g(\boldsymbol{\theta}) = \exp\left[-b(\boldsymbol{\theta})\right]$, $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{x}$ and $\boldsymbol{\phi}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ we have

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\left[\boldsymbol{\theta}^t\boldsymbol{x} - b(\boldsymbol{\theta})\right]$$

- Scalar random variable with a vector parameter:

$$\begin{aligned} p(x|\boldsymbol{\theta}) &= a(x)g(\boldsymbol{\theta})\exp\left[\sum_{k=1}^{K}\phi_k(\boldsymbol{\theta})h_k(x)\right] \\ &= a(x)g(\boldsymbol{\theta})\exp\left[\boldsymbol{\phi}^t(\boldsymbol{\theta})\boldsymbol{h}(x)\right] \end{aligned}$$

- Scalar random variable with a scalar parameter:

$$p(x|\theta) = \mathbf{Exf}(x|a,g,\phi,h) = a(x)g(\theta)\exp\left[\phi(\theta)h(x)\right]$$

$$p(x|\theta) = \theta\exp\left[-\theta x\right] = \exp\left[-\theta x + \ln\theta\right], \quad x \geq 0, \quad \theta \geq 0.$$

Exponential families and Conjugate priors

- A family $\mathcal{F}$ of probability distributions $\pi(\boldsymbol{\theta})$ on $\mathcal{T}$ is said to be conjugate (or closed under sampling) if, for every $\pi(\boldsymbol{\theta}) \in \mathcal{F}$, the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ also belongs to $\mathcal{F}$.
- Assume that $f(\boldsymbol{x}|\boldsymbol{\theta}) = l(\boldsymbol{\theta}|\boldsymbol{x}) = l(\boldsymbol{\theta}|\boldsymbol{t}(\boldsymbol{x}))$ where $\boldsymbol{t} = \{n, \boldsymbol{s}\} = \{n, s_1, \ldots, s_k\}$ is a vector of dimension $k+1$ and is sufficient statistics for $f(\boldsymbol{x}|\boldsymbol{\theta})$. Then, if there exists a vector $\{\tau_0, \boldsymbol{\tau}\} = \{\tau_0, \tau_1, \ldots, \tau_k\}$ such that

$$\pi(\boldsymbol{\theta}|\boldsymbol{\tau}) = \frac{f(\boldsymbol{s} = (\tau_1, \cdots, \tau_k)|\boldsymbol{\theta}, n = \tau_0)}{\int f(\boldsymbol{s} = (\tau_1, \cdots, \tau_k)|\boldsymbol{\theta}', n = \tau_0) \, \mathrm{d}\boldsymbol{\theta}'}$$

exists and defines a family $\mathcal{F}$ of distributions for $\boldsymbol{\theta} \in \mathcal{T}$, then the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{\tau})$ will remain in the same family $\mathcal{F}$. The prior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{\tau})$ is then a conjugate prior for the sampling distribution $f(\boldsymbol{x}|\boldsymbol{\theta})$.

Exponential families and Conjugate priors

- For a set of n i.i.d. samples $\{x_1, \cdots, x_n\}$ of a random variable $X \sim \mathbf{Exf}(x|a, g, \boldsymbol{\theta}, \boldsymbol{h})$ we have

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{n} f(x_j|\boldsymbol{\theta}) = [g(\boldsymbol{\theta})]^n \left(\prod_{j=1}^{n} a(x_j)\right) \exp\left[\sum_{k=1}^{K} \phi_k(\boldsymbol{\theta}) \sum_{j=1}^{n} h_k(x_j)\right]$$
$$= g^n(\boldsymbol{\theta})\, a(\boldsymbol{x}) \exp\left[\boldsymbol{\phi}^t(\boldsymbol{\theta}) \sum_{j=1}^{n} \boldsymbol{h}(x_j)\right],$$

where $a(\boldsymbol{x}) = \prod_{j=1}^{n} a(x_j)$. Then, using the factorization theorem it is easy to see that

$$\boldsymbol{t} = \left\{n, \sum_{j=1}^{n} h_1(x_j), \cdots, \sum_{j=1}^{n} h_K(x_j)\right\}$$

is a sufficient statistics for $\boldsymbol{\theta}$.

Exponential families and Conjugate priors

- A conjugate prior family for the exponential family

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = a(\boldsymbol{x})\, g(\boldsymbol{\theta}) \exp\left[\sum_{k=1}^{K} \phi_k(\boldsymbol{\theta})\, h_k(\boldsymbol{x})\right]$$

is given by

$$\pi(\boldsymbol{\theta}|\tau_0, \boldsymbol{\tau}) = z(\boldsymbol{\tau})[g(\boldsymbol{\theta})]^{\tau_0} \exp\left[\sum_{k=1}^{K} \tau_k \phi_k(\boldsymbol{\theta})\right]$$

- The associated posterior law is

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}, \tau_0, \boldsymbol{\tau}) \propto [g(\boldsymbol{\theta})]^{n+\tau_0} a(\boldsymbol{x}) z(\boldsymbol{\tau}) \exp\left[\sum_{k=1}^{K} \left(\tau_k + \sum_{j=1}^{n} h_k(x_j)\right) \phi_k\right.$$

Exponential families and Conjugate priors

- If

$$f(\boldsymbol{x}|\theta) = \textbf{Exfn}(\boldsymbol{x}|a(\boldsymbol{x}), g(\boldsymbol{\theta}), \phi, \boldsymbol{h}),$$

then a conjugate prior family is

$$\pi(\boldsymbol{\theta}|\boldsymbol{\tau}) = \textbf{Exfn}(\theta|g^{\tau_0}, z(\boldsymbol{\tau}), \boldsymbol{\tau}, \phi),$$

and the associated posterior law is

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{\tau}) = \textbf{Exfn}(\theta|g^{n+\tau_0}, a(\boldsymbol{x})\, z(\boldsymbol{\tau}), \boldsymbol{\tau}', \phi)$$

where

$$\tau_k' = \tau_k + \sum_{j=1}^{n} h_k(x_j)$$

or

$$\boldsymbol{\tau}' = \boldsymbol{\tau} + \bar{\boldsymbol{h}}, \quad \text{with} \quad \bar{h}_k = \sum_{j=1}^{n} h_k(x_j).$$

Exponential families and Conjugate priors

- If
$$f(\boldsymbol{x}|\theta) = a(\boldsymbol{x}) \exp\left[\boldsymbol{\theta}^t \boldsymbol{x} - b(\boldsymbol{\theta})\right]$$
Then a conjugate prior family is
$$\pi(\boldsymbol{\theta}|\boldsymbol{\tau}_0) = g(\boldsymbol{\theta}) \exp\left[\boldsymbol{\tau}_0^t \boldsymbol{\theta} - d(\boldsymbol{\tau}_0)\right]$$
and the corresponding posterior is
$$\pi(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{\tau}_0) = g(\boldsymbol{\theta}) \exp\left[\boldsymbol{\tau}_n^t \boldsymbol{\theta} - d(\boldsymbol{\tau}_n)\right] \quad \text{with} \quad \boldsymbol{\tau}_n = \boldsymbol{\tau}_0 + \bar{\boldsymbol{x}}$$
where
$$\bar{\boldsymbol{x}}_n = \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{x}_j$$

Exponential families and Conjugate priors

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = a(\boldsymbol{x}) \exp\left[\boldsymbol{\theta}^t \boldsymbol{x} - b(\boldsymbol{\theta})\right]$$

$$\pi(\boldsymbol{\theta}|\alpha_0, \boldsymbol{\tau}_0) = g(\alpha_0, \boldsymbol{\tau}_0) \exp\left[\alpha_0\, \boldsymbol{\tau}_0^t \boldsymbol{\theta} - \alpha_0 b(\boldsymbol{\tau}_0)\right]$$

$$\pi(\boldsymbol{\theta}|\alpha_0, \boldsymbol{\tau}_0, \boldsymbol{x}) = g(\alpha, \boldsymbol{\tau}) \exp\left[\alpha\, \boldsymbol{\tau}^t \boldsymbol{\theta} - \alpha b(\boldsymbol{\tau})\right]$$

$$\alpha = \alpha_0 + n \quad \text{and} \quad \boldsymbol{\tau} = \frac{\alpha_0 \boldsymbol{\tau}_0 + n\bar{\boldsymbol{x}}}{(\alpha_0 + n)})$$

$$\mathrm{E}\left\{\boldsymbol{X}|\boldsymbol{\theta}\right\} = \mathrm{E}\left\{\bar{\boldsymbol{X}}|\boldsymbol{\theta}\right\} = \nabla b(\boldsymbol{\theta})$$

$$\mathrm{E}\left\{\nabla b(\boldsymbol{\Theta})|\alpha_0, \boldsymbol{\tau}_0\right\} = \boldsymbol{\tau}_0$$

$$\mathrm{E}\left\{\nabla b(\boldsymbol{\theta})|\alpha_0, \boldsymbol{\tau}_0, \boldsymbol{x}\right\} = \frac{n\bar{\boldsymbol{x}} + \alpha_0 \boldsymbol{\tau}_0}{\alpha_0 + n} = \pi \bar{\boldsymbol{x}}_n + (1 - \pi)\boldsymbol{\tau}_0,$$

with $\pi = \frac{n}{\alpha_0 + n}$

Conjugate priors

Observation law $p(x\|\theta)$	Prior law $p(\theta\|\boldsymbol{\tau})$	Posterior law $p(\theta\|x,\boldsymbol{\tau}) \propto p(\theta\|\boldsymbol{\tau})p(x\|\theta)$
Binomial $\mathbf{Bin}(x\|n,\theta)$	Beta $\mathbf{Bet}(\theta\|\alpha,\beta)$	Beta $\mathbf{Bet}(\theta\|\alpha+x,\beta+n-x)$
Negative Binomial $\mathbf{NegBin}(x\|n,\theta)$	Beta $\mathbf{Bet}(\theta\|\alpha,\beta)$	Beta $\mathbf{Bet}(\theta\|\alpha+n,\beta+x)$
Multinomial $\mathbf{M}_k(x\|\theta_1,\cdots,\theta_k)$	Dirichlet $\mathbf{Di}_k(\theta\|\alpha_1,\cdots,\alpha_k)$	Dirichlet $\mathbf{Di}_k(\theta\|\alpha_1+x_1,\cdots,\alpha_k+x_k)$
Poisson $\mathbf{Pn}(x\|\theta)$	Gamma $\mathbf{Gam}(\theta\|\alpha,\beta)$	Gamma $\mathbf{Gam}(\theta\|\alpha+x,\beta+1)$

Conjugate priors

Observation law $p(x\|\theta)$	Prior law $p(\theta\|\boldsymbol{\tau})$	Posterior law $p(\theta\|x,\boldsymbol{\tau}) \propto p(\theta\|\boldsymbol{\tau})p($
Gamma $\mathbf{Gam}(x\|\nu,\theta)$	Gamma $\mathbf{Gam}(\theta\|\alpha,\beta)$	Gamma $\mathbf{Gam}(\theta\|\alpha+\nu,\beta+x)$
Beta $\mathbf{Bet}(x\|\alpha,\theta)$	Exponential $\mathbf{Ex}(\theta\|\lambda)$	Exponential $\mathbf{Ex}(\theta\|\lambda-\log(1-x))$
Normal $\mathbf{N}(x\|\theta,\sigma^2)$	Normal $\mathbf{N}(\theta\|\mu,\tau^2)$	Normal $\mathbf{N}\left(\mu\|\frac{\mu\sigma^2+\tau^2x}{\sigma^2+\tau^2},\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)$
Normal $\mathbf{N}(x\|\mu,1/\theta)$	Gamma $\mathbf{Gam}(\theta\|\alpha,\beta)$	Gamma $\mathbf{Gam}\left(\theta\|\alpha+\frac{1}{2},\beta+\frac{1}{2}\right.$
Normal $\mathbf{N}(x\|\theta,\theta^2)$	Generalized inverse Normal $\mathbf{INg}(\theta\|\alpha,\mu,\sigma)\propto$ $\|\theta\|^{-\alpha}\exp\left[-\frac{1}{2\sigma^2}\left(\frac{1}{\theta}-\mu\right)^2\right]$	Generalized inverse $\mathbf{INg}(\theta\|\alpha_n,\mu_n,\sigma_n)$