Financial Sig. Proc. & Mach. Learn.
Background Material

Danilo Mandic

room 813, ext: 46271

Department of Electrical and Electronic Engineering
Imperial College London, UK

d.mandic@imperial.ac.uk, URL: www.commsp.ee.ic.ac.uk/∼mandic

Probability vs. Statistics

For discrete RVs, $E\{X\} = \sum_{i=1}^{I} x_i P_X(x_i)$, where P_X is the probability function

Probability: A data modelling view, describes how data **will likely behave**

for example: $$average = E\{X\} = \int_{-\infty}^{\infty} x\, p_X(x)\, dx$$ no data here

Notice that there is no explicit mention of data here $\looparrowright$ x is a dummy variable and p_X is the pdf of a random variable X.

Statistics: A data analysis view, determines how data **did behave**

for example: $$average = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$$ no pdf here

Example: Consider N coarse-quantised data points, $x[0], \ldots, x[N-1]$. The signal has $M \ll N$ possible amplitude values, $V_1, \ldots, V_M$, with the corresponding relative frequencies, $N_1, \ldots, N_M$. Calculate the mean, $\bar{x}$.

Solution:

$$\bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \quad = \quad \frac{1}{N}\sum_{m=1}^{M} V_m N_m \quad = \quad \sum_{m=1}^{M} V_m \underbrace{\frac{N_m}{N}}_{\approx\, P(x=V_m)}$$

Probability vs. Statistics

(for discrete RVs, $E\{X\} = \sum_{i=1}^{I} x_i P_X(x_i)$, where P_X is the probability function)

Probability: A data modelling view, describes how data **will likely behave**

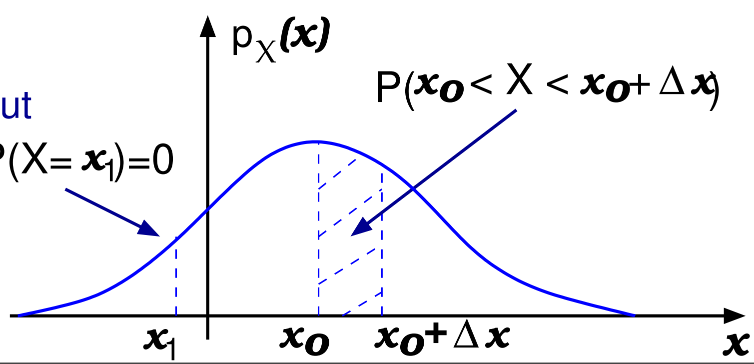
for example: $$average = E\{X\} = \int_{-\infty}^{\infty} x\, p_X(x)\, dx$$ no data here

Notice that there is no explicit mention of data here $\looparrowright$ x is a dummy variable and p_X is the pdf of a random variable X.

Statistics: A data analysis view, determines how data **did behave**

for example: $$average = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$$ no pdf here

Vagaries of probability: $P(x_0 < X < x_0 + \Delta x) = \int_{x_0}^{x_0+\Delta x} p_X(x)dx$



Notice that

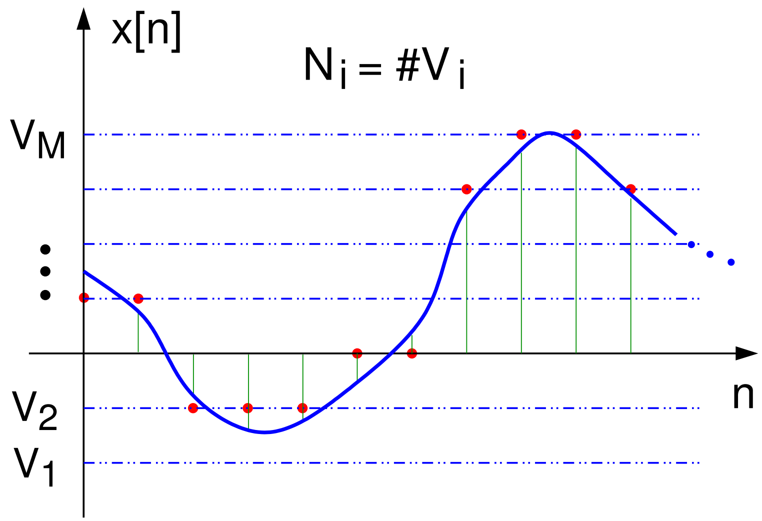
$$P(X = x_1) = 0$$

This appears odd, but otherwise the probabilities sum up to ∞

Statistics vs. Probability

Statistical inference $\looparrowright$ based on the observed data and supported by prob. theory

Vagaries of statistics: Consider N coarse-quantised data points, $x[0], \ldots, x[N-1]$. The quantised signal has $M \ll N$ possible amplitude values, $V_1, \ldots, V_M$, for which the corresponding relative frequencies are, $N_1 = \#V_1, \ldots, N_M = \#V_M$. Calculate the mean, $\bar{x}$.



Solution:

$$\bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \quad = \quad \frac{1}{N}\sum_{m=1}^{M} V_m N_m \quad = \quad \sum_{m=1}^{M} V_m \underbrace{\frac{N_m}{N}}_{\approx\ P(x=V_m)}$$

☞ Clearly, the factor $1/N$ does not imply "uniform distribution"

Statistical inference

Chinese for statistics is 统计 (summarizing & counting) and probability is 概率(论) ((theory of) randomness & chances),

Probability: Assumes perfect knowledge about the "population" of random data (through the pdf).

Typical question: There are 100 books on a bookshelf, 40 with red cover, 30 with blue cover, and 20 with green cover. What is the probability to randomly draw a blue book from the shelf?

Statistics: No knowledge about the types of books on the shelf, we need to infer properties about the "population" based on random samples of "objects" on the shelf $\looparrowright$ **statistical inference**.

Typical question: A random sampling of 20 books from the bookshelf produced X red books, Y blue books and Z green books. What is the total proportion of red, blue, and green books on the shelf?

Statistical inference is applied in many different contexts under the names of: data analysis, data mining, machine learning, classification, pattern recognition, clustering, regression, classification

Range of a matrix, span of a set of vectors

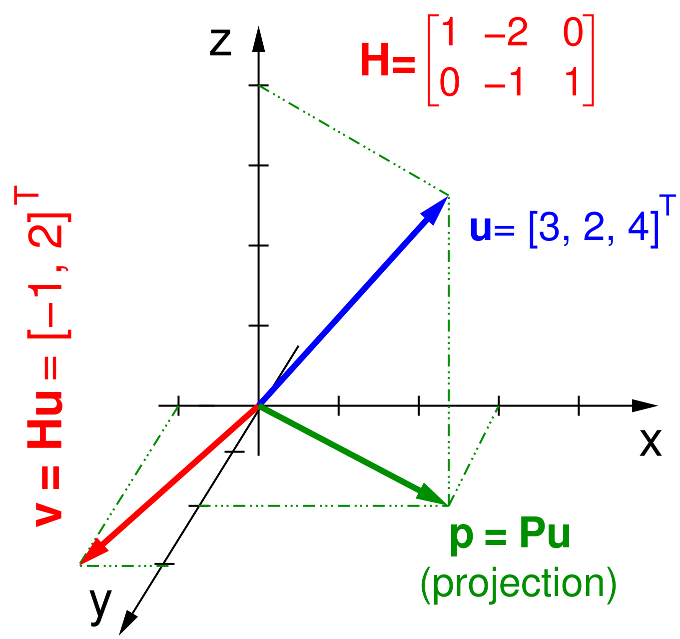
(a wide matrix transforms a vector space into another lower-dimensional one)

Consider a general 2×3 matrix $\mathbf{H}$ and a 3×1 vector $\mathbf{u}$

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \end{bmatrix} = [\mathbf{h}_1 \,|\, \mathbf{h}_2 \,|\, \mathbf{h}_3] \quad \text{where} \quad \mathbf{h}_i = \begin{bmatrix} h_{1i} \\ h_{2i} \end{bmatrix} \quad i = 1, 2, 3$$

Then,

$$\mathbf{v} = \mathbf{H}\,\mathbf{u} = [\mathbf{h}_1 \,|\, \mathbf{h}_2 \,|\, \mathbf{h}_3] \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = u_1\mathbf{h}_1 + u_2\mathbf{h}_2 + u_3\mathbf{h}_3 \;\in \mathbb{R}^{2\times 1}$$



Example: $\mathbf{H} \in \mathbb{R}^{2\times 3}, \mathbf{u} \in \mathbb{R}^{3\times 1}$

∘ Clearly, $\mathbf{v}$ is a linear combination of the columns of the matrix $\mathbf{H}$, $\mathbf{h}_i \in \mathbb{R}^{2\times 1}$

∘ Vector $\mathbf{v} = [-1, 2]^T$ therefore lies in the span of the columns of $\mathbf{H}$, i.e. in $\mathbb{R}^2$

☞ This dimensionality reduction is not a projection $\mathbf{p} = \mathbf{P}\mathbf{u}$, where $\mathbf{P} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$

Quadratic forms and positive–(semi)definite matrices

Quadratic forms appear often in data analysis, and are expressed as

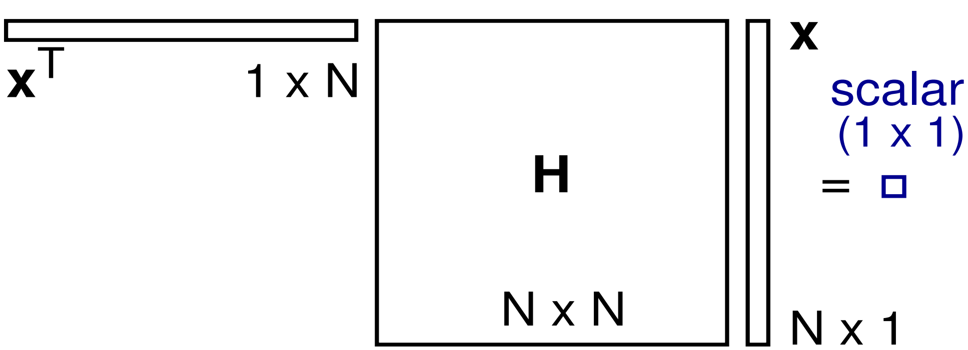
$$\mathbf{x}^T\mathbf{H}\mathbf{x} \qquad\qquad \mathbf{x}\in\mathbb{R}^{N\times 1},\ \mathbf{H}\in\mathbb{R}^{N\times N}$$

For simplicity, consider a 2nd order case, where

$$\underset{\uparrow\ variable\ \ vector}{\mathbf{x}=\begin{bmatrix} x_1\\ x_2\end{bmatrix}} \qquad\qquad \underset{\uparrow\ fixed\ \ matrix}{\mathbf{H}=\begin{bmatrix} h_{11} & h_{12}\\ h_{21} & h_{22}\end{bmatrix}}$$

The quadratic form $Q_{\mathbf{H}}(\mathbf{x}) = Q_{\mathbf{H}}(x_1,x_2)$ of a matrix $\mathbf{H}$ is a scalar given by

$$Q_{\mathbf{H}}(x_1,x_2) = \mathbf{x}^T\,\mathbf{H}\,\mathbf{x} = \sum_{i=1}^{2}\sum_{j=1}^{2} h_{ij}x_ix_j = h_{11}x_1^2 + h_{22}x_2^2 + (h_{12}+h_{21})x_1x_2$$



◦ If $Q_{\mathbf{H}}(\mathbf{x})\geq 0$, for any $\mathbf{x}\neq\mathbf{0}$ then the matrix $\mathbf{H}$ is called positive semidefinite $(\mathbf{H}\geq\mathbf{0})$

◦ The matrix $\mathbf{H}$ is positive definite if $Q_{\mathbf{H}}(\mathbf{x})>0, \forall\mathbf{x}\neq\mathbf{0}$

Whitening operation of the inverse covariance matrix

Consider a general linear model, given by ($\mathbf{H}$ must be full rank)

$$\textit{known observation matrix} \downarrow \quad \downarrow \textit{unknown parameters to be estimated}$$

$$\mathbf{x} = \mathbf{H}\,\boldsymbol{\theta} + \mathbf{w}$$

$$\textit{observed data} \uparrow \qquad\qquad \uparrow \textit{noise} \sim \mathcal{N}(0, \mathbf{C})$$

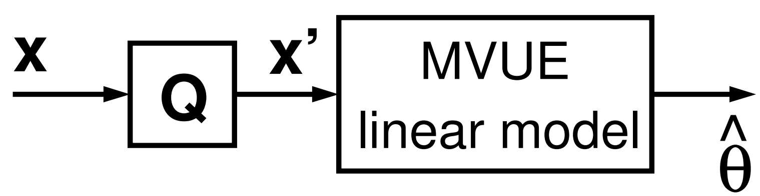
In many situations (CRLB, multivar. Gaussian), $\mathbf{C}$ needs to be invertible.

Then for $\mathbf{C} \geq \mathbf{0}$, $\exists$ an invertible $\mathbf{Q}$ $s.t.$ $\mathbf{C}^{-1} = \mathbf{Q}^T\mathbf{Q} \implies \mathbf{C} = \mathbf{Q}^{-1}(\mathbf{Q}^T)^{-1}$

We can now transform (rotate) data $\mathbf{x}$ as $\mathbf{x}' = \mathbf{Q}\mathbf{x}$, to give

$$\mathbf{x}' = \mathbf{Q}\,\mathbf{x} = \mathbf{Q}\,\mathbf{H}\,\boldsymbol{\theta} + \mathbf{Q}\,\mathbf{w} = \mathbf{H}'\,\boldsymbol{\theta} + \mathbf{w}'$$

$$\begin{aligned} E\{\mathbf{w}'\mathbf{w}'^T\} &= E\{\mathbf{Q}\mathbf{w}\,(\mathbf{Q}\mathbf{w})^T\} = E\{\mathbf{Q}\mathbf{w}\,\mathbf{w}^T\mathbf{Q}^T\} \\ &= \mathbf{Q}\,\mathbf{C}\,\mathbf{Q}^T = \mathbf{Q}\,\underbrace{\mathbf{Q}^{-1}(\mathbf{Q}^T)^{-1}}_{\mathbf{C}}\,\mathbf{Q}^T = \mathbf{I} \end{aligned}$$



☞ **$\mathbf{w}' = \mathbf{Q}\mathbf{w}$ is now white!**

More on quadratic forms and covariance matrices

Consider a vector of random variables $\mathbf{x} = [X_0, \ldots, X_{N-1}]^T \in \mathbb{R}^{N \times 1}$. Then, if these random variables are jointly Gaussian, their PDF is given by

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{det(\mathbf{C})}} \, e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x}-\boldsymbol{\mu})} \quad \leftarrow \text{quadratic form}$$

where $\boldsymbol{\mu} = E\{\mathbf{x}\}$ is mean vec. and $\mathbf{C} = E\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\}$ covar. mat..

For two jointly Gaussian random variables X_1 and X_2, the means $\mu_1 = E\{X_1\}, \mu_2 = E\{X_2\}$, variances $\sigma_1^2 = var(X_1), \sigma_2^2 = var(X_2)$, co–variance $\sigma_{12} = E\{(X_1 - \mu_1)(X_2 - \mu_2)\}$, and the correlation coefficient $\rho = \sigma_{12}/(\sigma_1\sigma_2)$.

Then,

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \, e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]}$$

Obviously, if X_1 and X_2 are uncorrelated, then $p(x_1, x_2) = p(x_1)p(x_2)$, and

$$p(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \, \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}}$$

Quadratic forms, covariance matrices, and Gaussian PDF

For convenience, assume zero–mean data a $\mathbf{x} = [x_1, x_2]^T \in \mathbb{R}^{2\times 1}$

$$\text{Then,} \qquad p(\mathbf{x}) = \frac{1}{2\pi\sqrt{det(\mathbf{C})}} exp\Big[-\frac{1}{2} \underbrace{\mathbf{x}^T \ \mathbf{C}^{-1} \ \mathbf{x}}_{\text{quad. form, scalar}} \Big]$$

This is a quadratic form, as we can write $\mathbf{C}^{-1} = \mathbf{A}$, another matrix. The "equi–potential" contours of this PDF are then determined by

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = k \qquad \text{k is a constant}$$

For this 2D case, the equi–potential contours of $\mathbf{x}^T \mathbf{A} \mathbf{x}$ are given by

$$a_{11}x_1^2 + a_{22}x_2^2 + (a_{12} + a_{21})x_1 x_2 = k \qquad \leftarrow \text{equation of an ellipse}$$

Because $\mathbf{C}$ is symmetric, $\mathbf{C}^{-1}$ is symmetric too, so that $a_{12} = a_{21}$

Uncorrelated x_1 and x_2

$$\mathbf{C}^{-1} = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix} \Rightarrow \mathbf{C} = \begin{bmatrix} \sigma_1^2 = \frac{1}{a_{11}} & 0 \\ 0 & \sigma_2^2 = \frac{1}{a_{22}} \end{bmatrix}$$

Ellipse aligned with the axes, since $a_{12} = a_{21} = 0$.

Correlated x_1 and x_2

$$\mathbf{C} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}, \sigma_{12} = \sigma_{21}$$

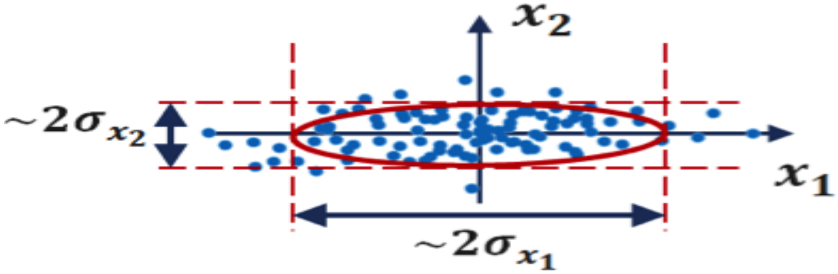
Ellipse not aligned with the axes.

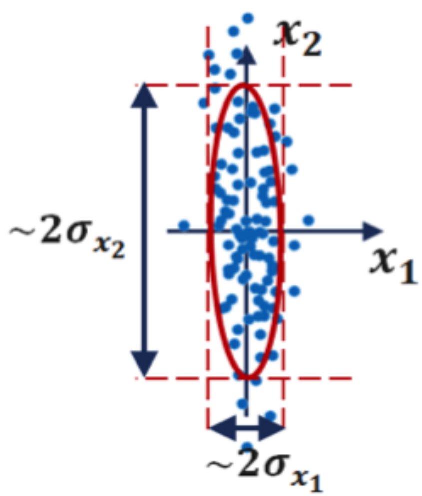
Correlation between RVs and (error) ellipsoids

Consider a bivariate quadratic form, $\mathbf{x}^T \mathbf{C} \mathbf{x} = k$ (equi-potential ellipses):

Off-diagonal elements $= 0$

x_1 and x_2 are uncorrelated, $\sigma_1^2 > \sigma_2^2$

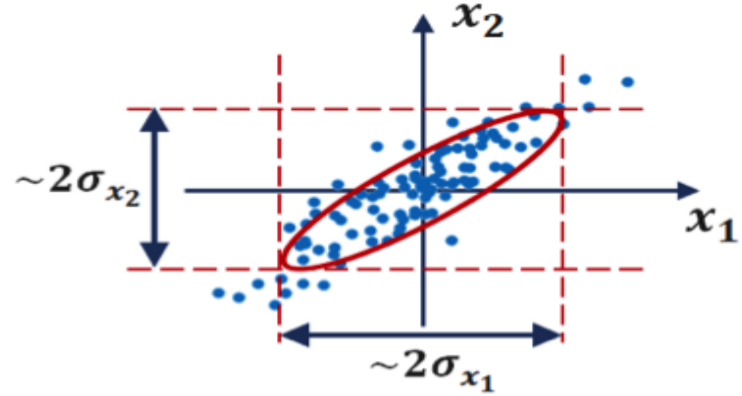


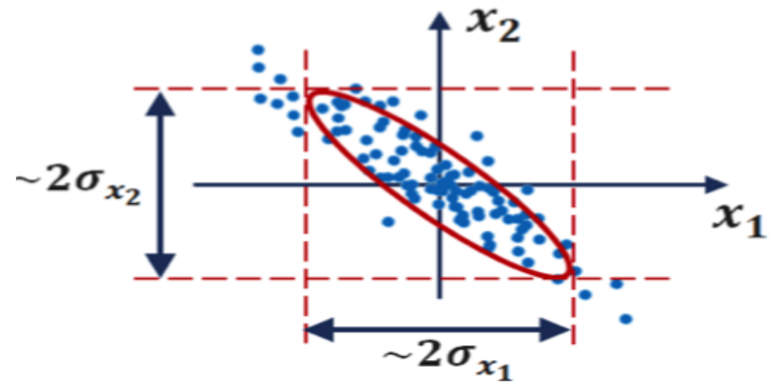


x_1 and x_2 are uncorrelated, $\sigma_1^2 < \sigma_2^2$

Off-diagonal elements $\neq 0$

x_1 and x_2 are positively correlated



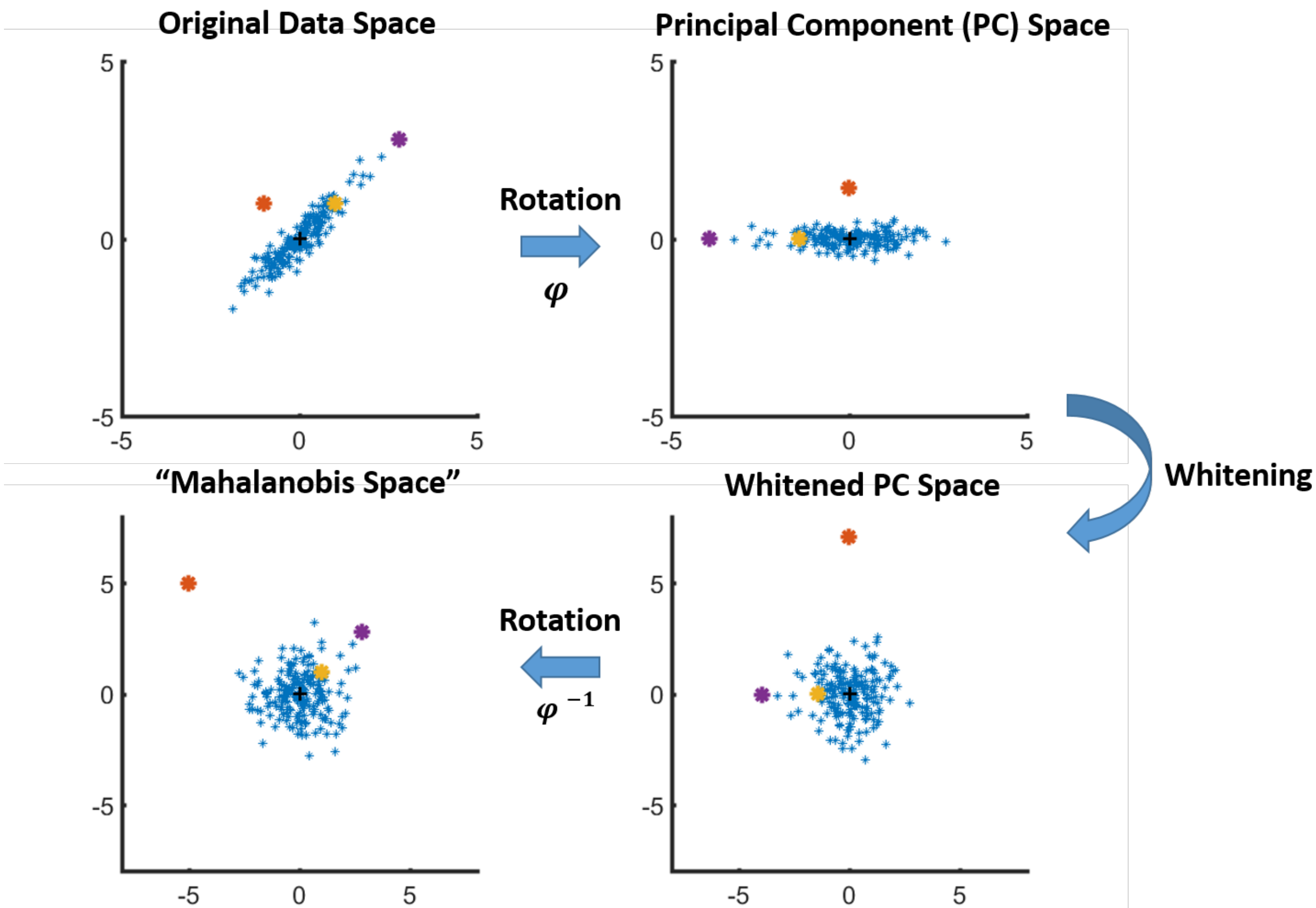


x_1 and x_2 are negatively correlated

Mahalanobis distance

Euclidean distance : $\| \mathbf{x}_1 - \mathbf{x}_2 \| = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^T (\mathbf{x}_1 - \mathbf{x}_2)}$

Mahalanobis distance : $\| \mathbf{x}_1 - \mathbf{x}_2 \|_M = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^T \mathbf{C}^{-1} (\mathbf{x}_1 - \mathbf{x}_2)}$

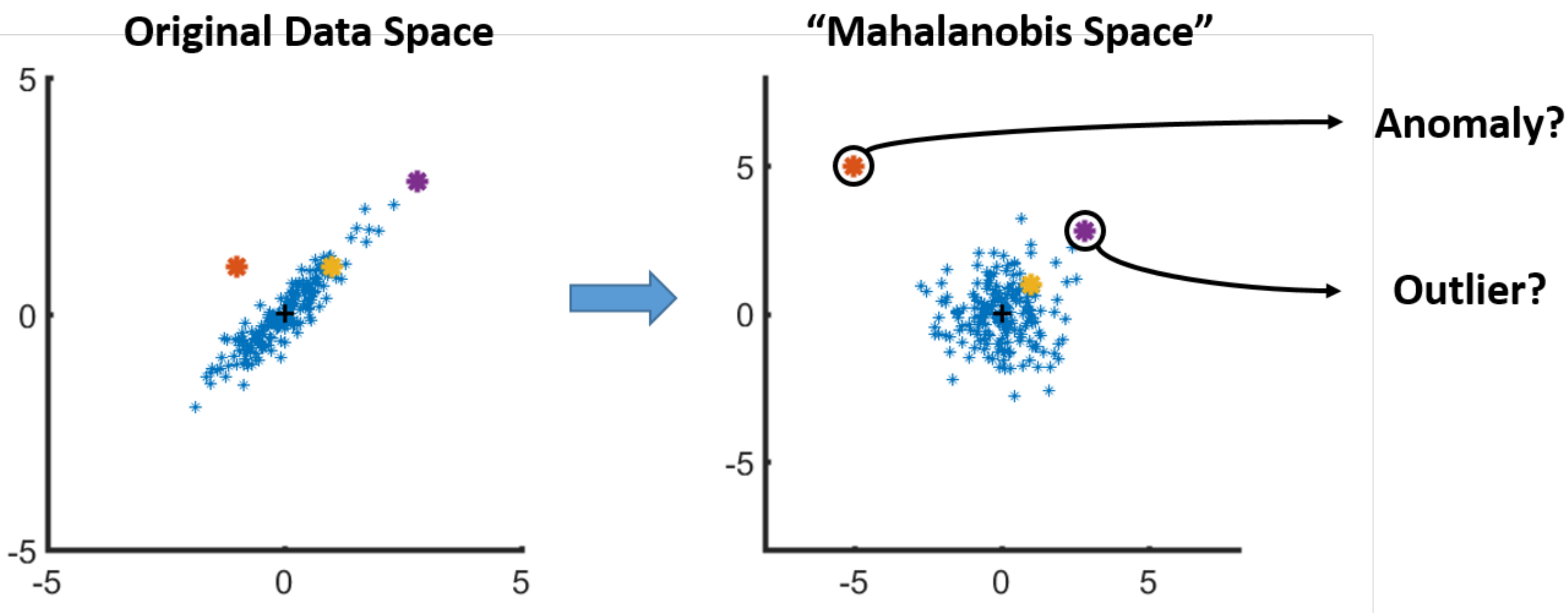


Advantages of the Mahalanobis distance

◦ Mahalanobis is scale–free and provides likelihood whether the data are generated by the same signal generating system (do data points belong to the same distribution)

◦ It is very useful at distinguishing between the outliers and anomalies

◦ Limitations ↬ from the PCA backbone (# data, normality, linearity)



Random variable (RV), some general observations

Random variable $\looparrowright$ quantifies the outcome of a random event.

For example, "heads" or "tails" on a coin or a blue square on Rubik's cube are not random variables per se, but can be made random variables *through numerical characterisation*.

☞ We therefore do not know how to determine the value of a RV, but can specify the probability of occurrence of a certain value of a RV.

A random var. X with the pdf

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

is called a Gaussian RV.

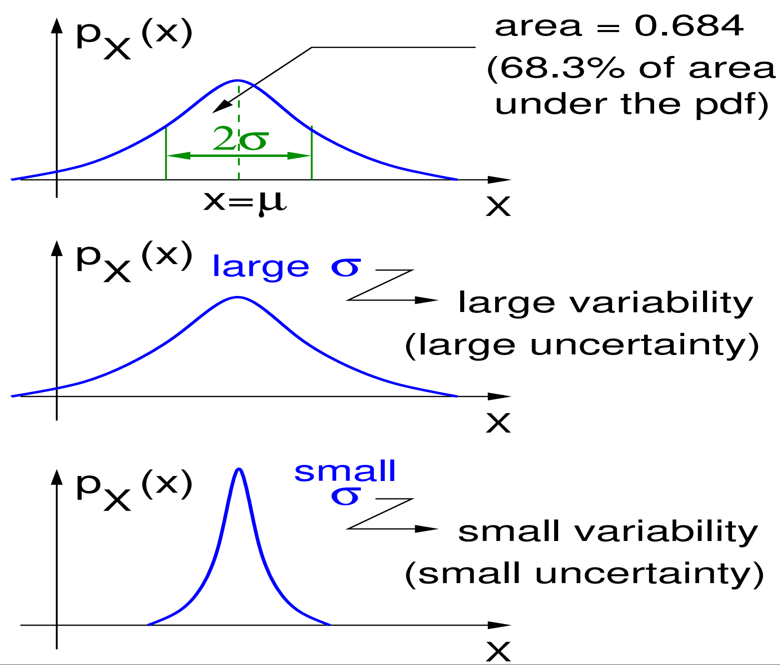
$\rightarrow$ μ is the mean of a RV X

$\rightarrow$ σ is the standard deviation of a RV X, and $\sigma > 0$

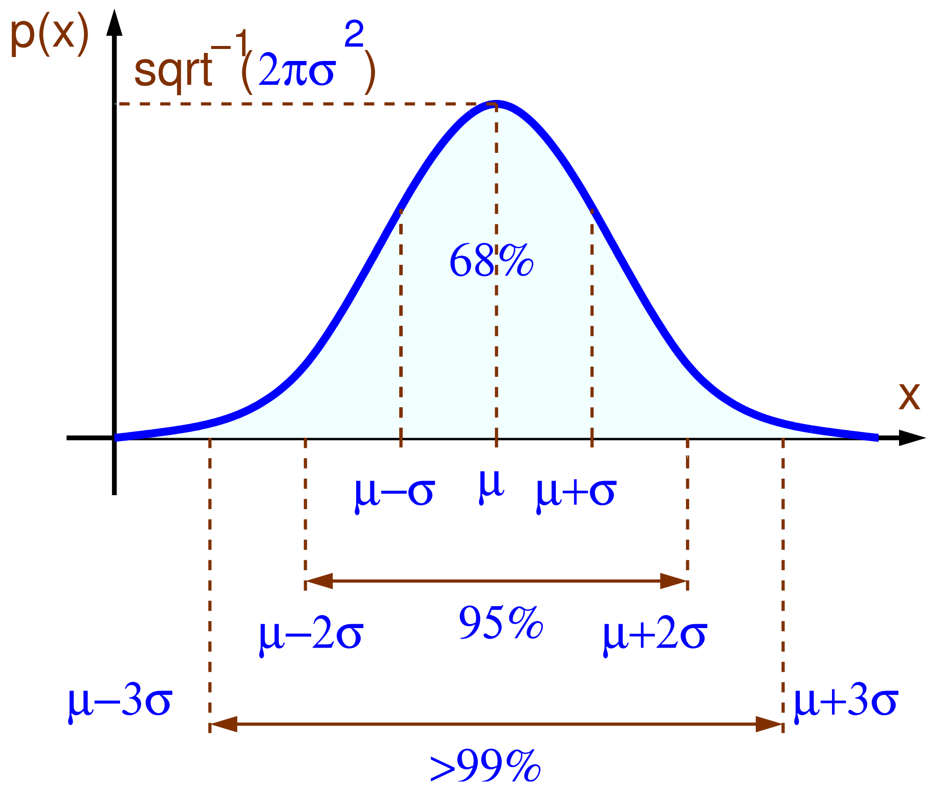
$\rightarrow$ σ^2 is the variance of a RV X

So, we can write $X \sim \mathcal{N}(\mu, \sigma^2)$

Variance effect on Gaussian pdf



Properties of the Gaussian distribution



1) If x and y are jointly Gaussian, then for any constants a and b the random variable

$$z = ax + by$$

is Gaussian with mean

$$m_z = am_x + bm_y$$

and variance

$$\sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\sigma_x\sigma_y\rho_{xy}$$

2) If two jointly Gaussian random variables are *uncorrelated* ($\rho_{xy} = 0$) then they are statistically independent,

$$f_{x,y} = f(x)f(y)$$

For $\mu = 0$, $\sigma = 1$, the inflection points are ± 1

Conditional pdf
“slice and normalise” the joint pdf $p(x,y)$

Formal definition

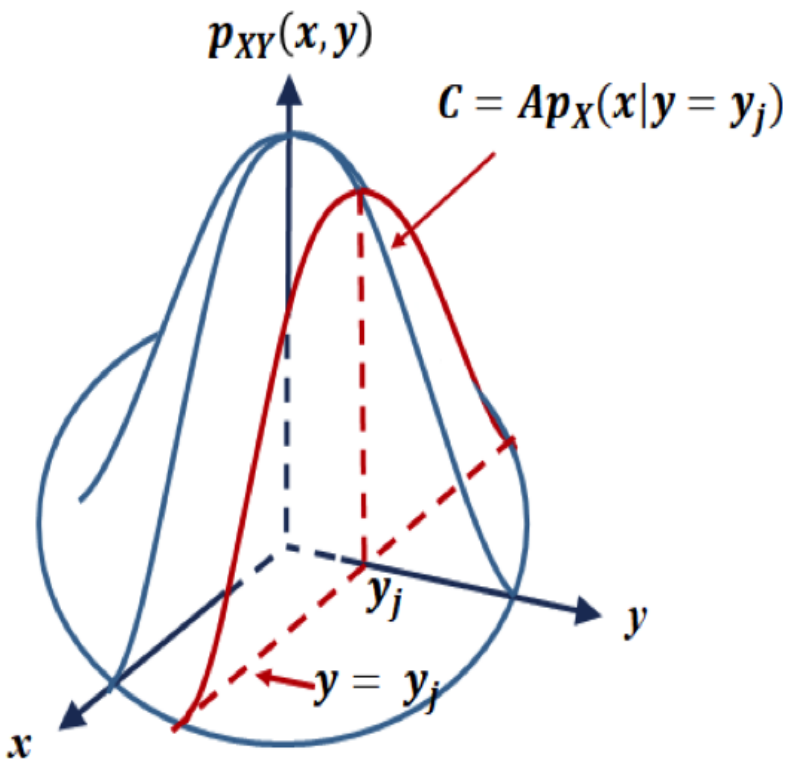
$$p_{Y|X}(y|x) = \begin{cases} \frac{p_{XY}(x,y)}{p_X(x)}, & p_X(x) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

x is held fixed

or more often

$$p(x|y) = \begin{cases} \frac{p(x,y)}{p(y)}, & p(y) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

y is held fixed



Conditional pdf $p(x|y)$

Depends on joint pdf $p(x,y)$ because there are two rand. variables, x and y.

Example: Length of holidays, X, conditioned on the salary $Y =$£60k?

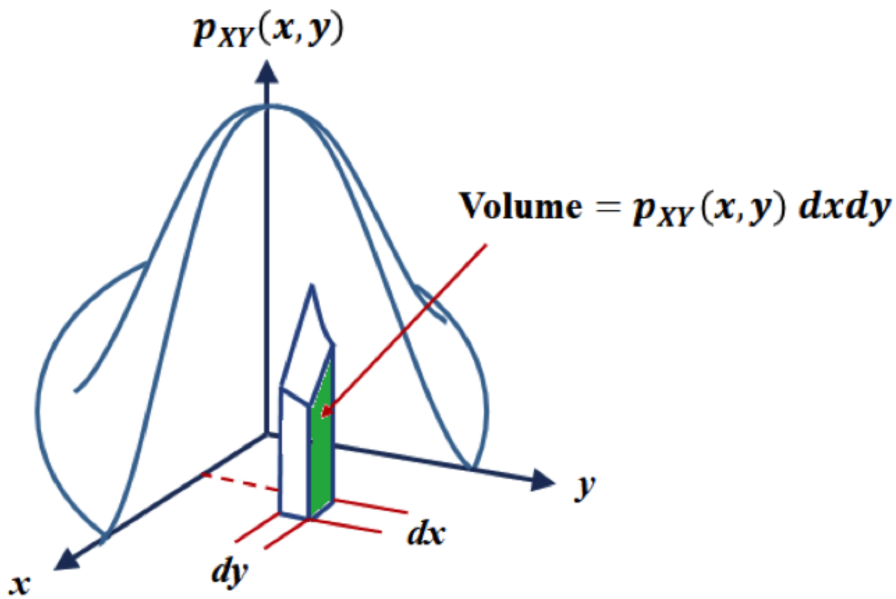
Ans: Find all people who make exactly £60k, how is holiday length distributed?

We therefore:

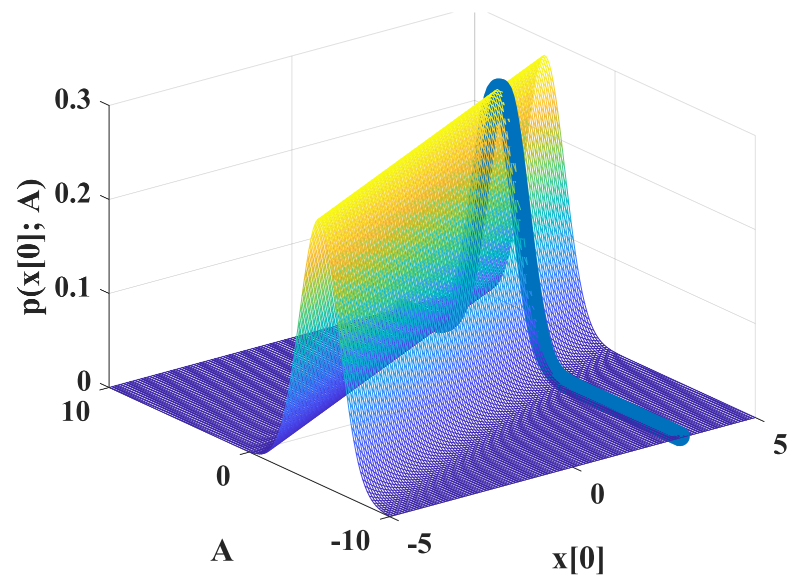
- **slice** the joint $p(x,y)$ at $Y = \pounds 60k$
- **normalise** by $p_Y(60,000)$ so that $p(x|y)=p(x,60k)/p_Y(60k)$ is valid pdf

Joint pdf vs. conditional pdf vs. parametrised pdf

Joint pdf $p(x, y)$



Parametrised pdf $p(x[0]; A)$

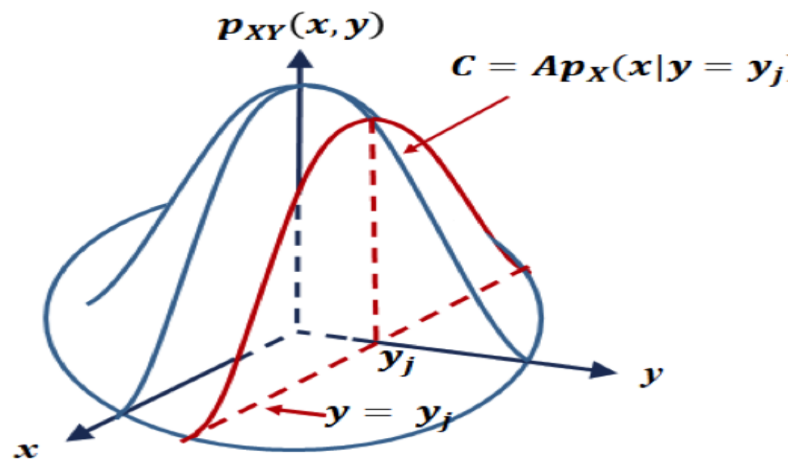


○ The joint pdf $p(x, y)$ is a truly 2D function of the rand. variables X and Y

○ The parametrised $p(x[0]; A)$ should be looked at as a function of A for a fixed value of observed data $x[0]$ (1-D)

The conditional pdf $\longrightarrow$

a 1-D funct. (slice and normalise).



Independent random variables

Independence: Contour ellipses are aligned with the x or y axes (Gaussian RVs)

Independent random variables: Neither random variable impacts the other one statistically.

For example, Argentinian debt and maximum tide in the Thames.

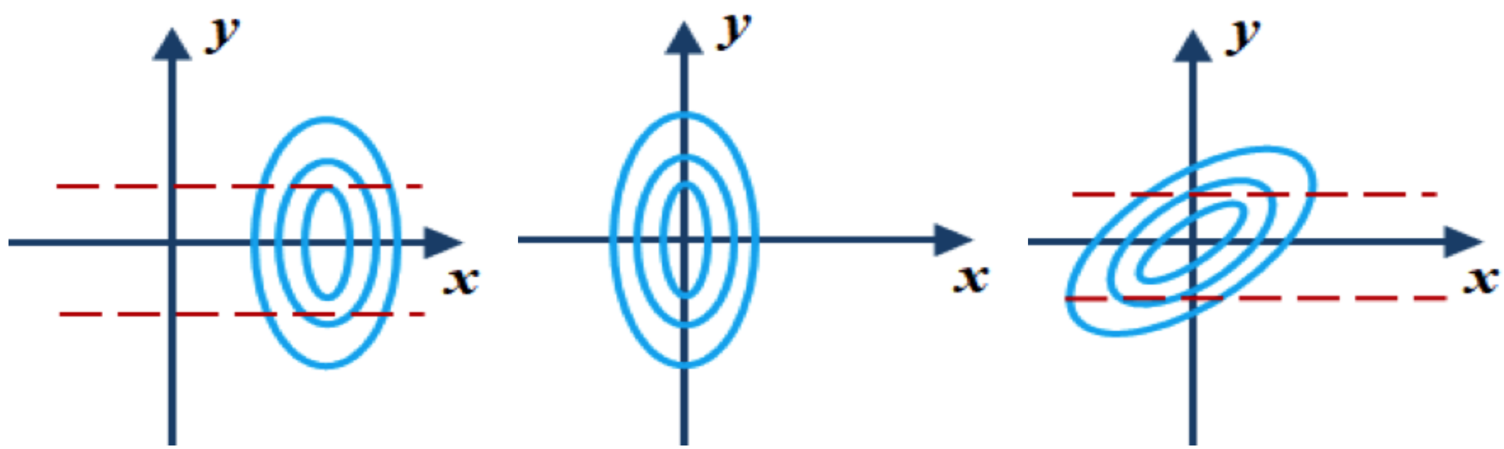
☞ For independent RVs, conditioning does not change the original PDF (the values Y will likely take are irrelevant to the value X has already taken).

$$p_{Y|X=x}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)} = p_Y(y) \qquad p_{X|Y=y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)} = p_X(x)$$

Independent non–zero mean

Independent zero–mean

Dependent (different slices $\nrightarrow$ different curves)



Confusing notation, X vs. x for a random variable

First, let us re–state the "independent" RV s result

$$p_{XY}(x,y) = p_X(x)\,p_Y(y) \quad \text{then} \quad p_{Y|X=x}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)} = \frac{p_X(x)\,p_Y(y)}{p_X(x)} = p_Y(y)$$

A pdf tells us everything about a RV. Other useful descriptors of a RV:

- ∘ The Mean $\looparrowright$ describes the centroid of the pdf
- ∘ The Variance of a RV $\looparrowright$ describes the spread of the pdf
- ∘ Correlation of two or more RVs $\looparrowright$ describes the "tilt" of the joint pdf

Theoretical view of Mean and Variance:

discrete RVs

$$\underbrace{E\{X\}}_{\text{random variable X}} = \sum_{i=1}^{I} x_i \underbrace{P_X(x_i)}_{\text{probability function}}$$

continuous RVs

$$E\{X\} = \int_{-\infty}^{\infty} \underbrace{x}_{\text{dummy variable}} \underbrace{p_X(x)}_{\text{pdf of RV X}} dx$$

Variance: $\quad \sigma^2 = \sigma_X^2 = E\{(X-\mu_x)^2\} = \int (x-\mu_x)^2 p_X(x)dx$

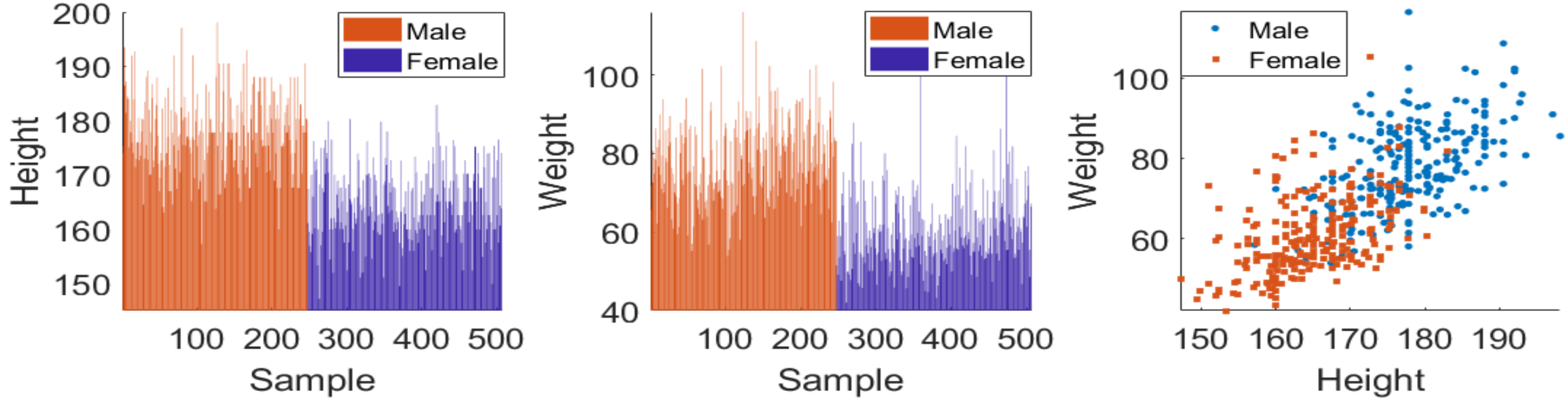
For a zero–mean random variable, $\sigma^2 = E\{X^2\} = \int x^2 p_X(x)dx$

Correlation: Data analysis view

Consider two random variables:

- ◦ X which represents the height of people, and
- ◦ Y which represents the weight of people

Then, clearly the random variables X and Y are correlated, as shown in the following scatter diagram



These two variables are obviously **positively correlated**, as on the average taller people will also be heavier.

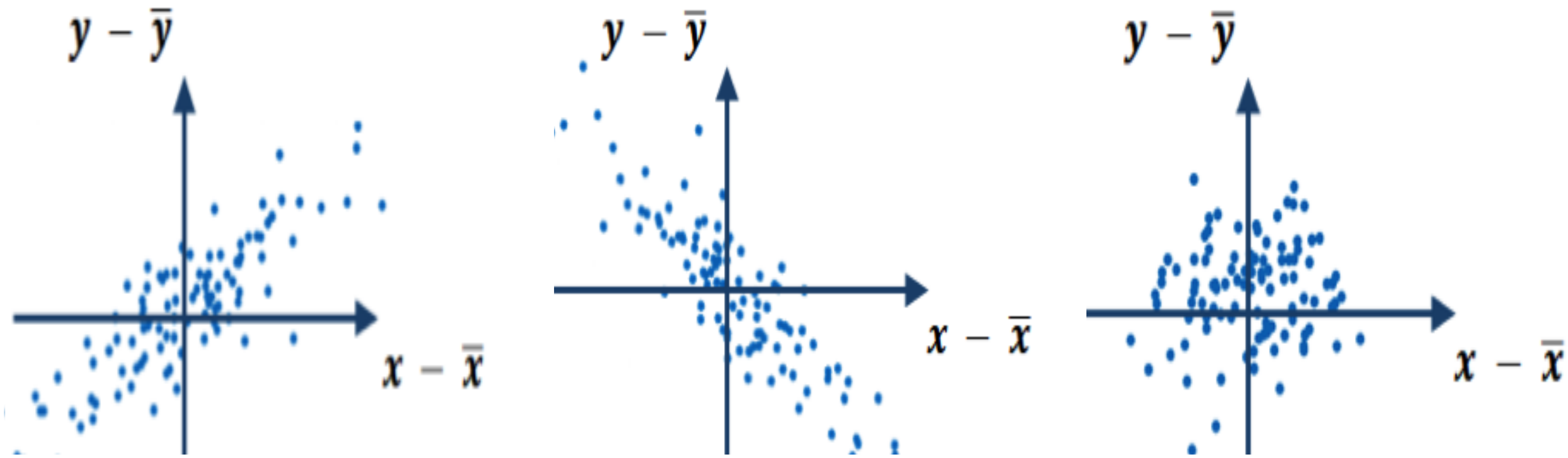
An outlier: A perfectly good data point which occurs rarely, that is, it is in the tails of the distribution.

Anomaly: A data point(s) which comes from a different process/pdf.

Covariance: A data analysis view

Covariance = Correlation of zero–mean variables

$$c_{xy} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_x)\,(y_n - \mu_y)$$



Positive correlation	Negative correlation	Zero correlation (uncorrel.)
height & weight	height & comfort in economy cl.	height & eye colour
"best friends"	**"worst enemies"**	**"complete strangers"**

These scatter diagrams are very informative, e.g. *complex noncircularity*

Correlation: Probability view

Correlation: $r_{XY} = E\{XY\}$ Covariance: $\sigma_{XY} = E\{(X-\mu_x)(Y-\mu_y)\}$

more often "cross-correlation"

original random var. (non–centred RVs)

more often "cross-covariance"

centered RVs (zero mean)

$$\sigma_{XY} = \int\int (x-\mu_x)(y-\mu_y)\, p_{XY}(x,y)dxdy \quad = r_{XY} \quad \text{for zero-mean} \quad X, Y$$

Auto-correlation: $r_{XX} = E\{XX\}$ (Auto-Co)Variance: $\sigma^2 = \sigma_{XX} = E\{(X-\mu_x)(X-\mu_x)\}$

Uncorrelated RVs: $\sigma_{XY} = E\{(X-\mu_x)(Y-\mu_y)\} = 0 \quad \Rightarrow \quad \mathbf{r_{XY} = E\{XY\}} = \mu_\mathbf{x}\mu_\mathbf{y}$

for uncorrelated random variables $\underbrace{E\{X\}E\{Y\} = \mu_x\mu_y}_{\text{separability of means}}$

Uncorrelated RVs $\nleftrightarrow$ zero cross-covariance $\sigma_{XY} \Leftrightarrow$ cross–correlation $= \mu_x\mu_y$

☞ X and Y are independent $\Rightarrow$ $p_{XY}(x,y) = p_X(x)p_Y(y)$ (separable joint pdf)

☞ X and Y are uncorrelated $\Rightarrow$ $E\{XY\} = E\{X\}E\{Y\}$

More confusing terminology and linear transforms of Gaussians

Recall that

Correlation: $r_{XY} = E\{XY\}$ Covariance: $\sigma_{XY} = E\big\{(X-\mu_x)(Y-\mu_y)\big\}$

Then, the **correlation coefficient** is defined as

$$\rho = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad -1 \le \rho \le 1$$

that is, based on the covariance σ_{XY}. For zero mean RVs, $r_{XY} = \sigma_{XY}$.

Linear transforms of multivariate Gaussian RVs. Consider a vector $\mathbf{x} = [X_1, \ldots, X_N]$ of jointly Gaussian random variables (multivariate Gaussian), with the covariance matrix $\mathbf{C}_{xx}$ and the mean vector $\boldsymbol{\mu}_x$.

Then, the linear transform $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ is also jointly Gaussian

For the so produced random vector $\mathbf{y}$,

$$\boldsymbol{\mu}_y = E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\mu}_x + \mathbf{b} \qquad \mathbf{C}_y = E\{(\mathbf{y}-\boldsymbol{\mu}_y)(\mathbf{y}-\boldsymbol{\mu}_y)^T\} = \mathbf{A}\mathbf{C}_{xx}\mathbf{A}^T$$

☞ The model is general, and gives a "sum of Gaussians" for $\mathbf{A} = array(1)$.

Moments of Gaussian random var. and Chi-squared distr.
Two very useful results

For a zero–mean Gaussian random variable $X \sim \mathcal{N}(0, \sigma^2)$

$$E\{X^n\} = \begin{cases} 1 \times 3 \times 5 \times \cdots \times (n-1)\sigma^n, & n \text{ even} \\ 0, & n \text{ odd} \end{cases}$$

Fourth order moment of Gaussians. The fourth–order moment of jointly Gaussian zero–mean random variables, X_1, X_2, X_3, X_4, is factorised as

$$E\{X_1X_2X_3X_4\} = E\{X_1X_2\}E\{X_3X_4\} + E\{X_1X_3\}E\{X_2X_4\} + E\{X_1X_4\}E\{X_2X_3\}$$

◦ This can also be applied to $E\{X^2Y^2\}$ for jointly Gaussian X and Y

Chi-squared χ^2 random variable Y based on zero–mean independent and jointly Gaussian RVs, $X_1, \ldots, X_N$, is given by

$$Y = X_1^2 + X_2^2 + \cdots X_N^2$$

with the probability density function

$$p(y) = \begin{cases} \frac{1}{2^{N/2}\Gamma(N/2)}\, y^{N/2-1}\, e^{-y/2}, & \text{for } y \geq 0 \\ 0, & \text{for } y < 0 \end{cases}$$

where Γ is the gamma distr. We then have $E\{Y\} = N$ and $var(Y) = 2N$.

Correlation and covariance matrices

The correlation matrix of random vector $\mathbf{x} = [x_0, \ldots, x_{N-1}]^T$ is given by

$$\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^T\} = \begin{bmatrix} E\{x_0x_0\} & E\{x_0x_1\} & \cdots & E\{x_0x_{N-1}\} \\ E\{x_1x_0\} & E\{x_1x_1\} & \cdots & E\{x_1x_{N-1}\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{x_{N-1}x_0\} & E\{x_{N-1}x_1\} & \cdots & E\{x_{N-1}x_{N-1}\} \end{bmatrix} = \begin{bmatrix} r_0 & r_1 & \cdots & r_{N-1} \\ r_1 & r_0 & \cdots & r_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N-1} & r_{N-2} & \cdots & r_0 \end{bmatrix}$$

and is symmetric, positive-semidefinite, Toeplitz, and of rank-N.

The covariance matrix, $\mathbf{C}_{xx} = E\{(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^T\}$, is given by

$$\mathbf{C}_{xx} = \begin{bmatrix} c(0) & c(1) & \cdots & c(N-1) \\ c(1) & c(0) & \cdots & c(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ c(N-1) & c(N-2) & \cdots & c(0) \end{bmatrix}$$

- ☞ Then, $\mathbf{C}_{xx} = \mathbf{R}_{xx} - \boldsymbol{\mu}_x\boldsymbol{\mu}_x^T$. For zero–mean random vectors, $\mathbf{C}_{xx} = \mathbf{R}_{xx}$.
- ☞ In probability theory, $\mathbf{C}_{xx}$ is also known the **dispersion matrix**, e.g. in the context of Gaussian RVs.

From a univariate to a multivariate Gaussian distribution

If each of the L samples of a random signal $x[i], i = 1, 2, \dots, L$ is Gaussian distributed, then

$$p(x[i]) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\, e^{\frac{\left(x[i]-\mu(i)\right)^2}{2\sigma_i^2}} \qquad i = 0, \dots, L-1$$

This distribution is denoted by $\mathcal{N}(\mu(i), \sigma_i^2)$.

The joint pdf of L samples $x[n_0], x[n_1], \dots, x[n_{L-1}]$ is then

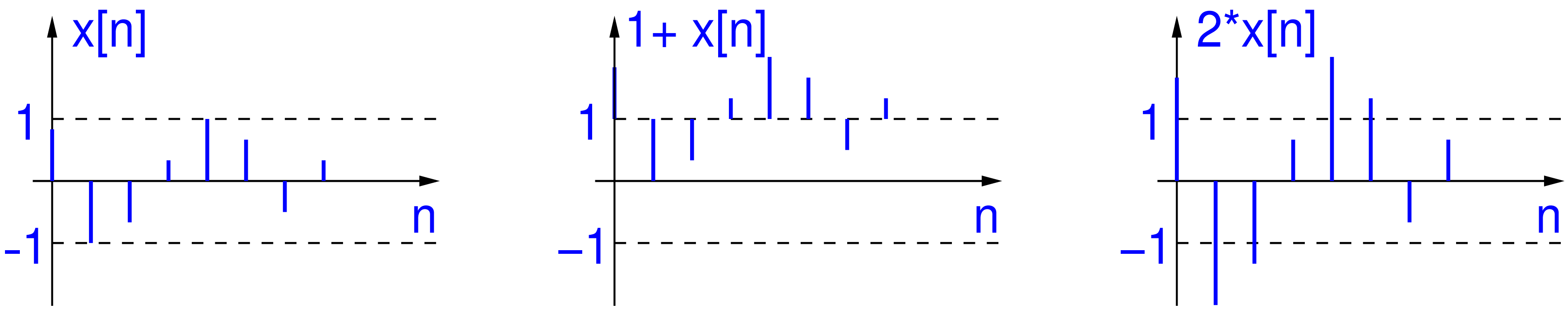
$$\begin{aligned} p(\mathbf{x}) &= p(x[n_0], x[n_1], \dots, x[n_{L-1}]) \\ p(\mathbf{x}) &= \frac{1}{[2\pi]^{L/2} det(\mathbf{C})^{1/2}} e^{\frac{(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}} = \frac{1}{(2\pi\sigma^2)^{L/2}} e^{\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{L-1}(x[n]-\mu)^2\right]} \end{aligned}$$

where $\mathbf{x} = [x[n_0], x[n_1], \dots, x[n_{L-1}]]$, $\boldsymbol{\mu} = [\mu[n_0], \mu[n_1], \dots, \mu[n_{L-1}]]$ and $\mathbf{C}$ is a covariance matrix with determinant Δ.

Some properties of the statistical Expectation Operator

$$E\{X+Y\} = E\{X\} + E\{Y\}, \quad E\{aX\} = aE\{X\}, \quad E\{g(X)\} = \int g(x)p_X(x)dx$$

$$var(X+Y) = \begin{cases} \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}, & X,Y \text{ correlated} \\ \sigma_X^2 + \sigma_Y^2, & X,Y \text{ uncorrelated} \end{cases} \qquad \begin{aligned} var(aX) &= a^2\sigma_X^2 \\ var(a+X) &= \sigma_X^2 \end{aligned}$$



This is because

$$\begin{aligned} var(X+Y) &= E\{(X+Y-\mu_x-\mu_y)^2\} = E\{(X_c+Y_c)^2\} \qquad X_c = X - \mu_x \\ &= E\{X_c^2 + Y_c^2 + 2X_cY_c\} = E\{X_c^2\} + E\{Y_c^2\} + 2E\{X_cY_c\} \\ &= \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} \end{aligned}$$

Notes:

∘