# BME I5100: Biomedical Signal Processing 

## Jointly Distributed Random Variables



Lucas C. Parra

Biomedical Engineering Department City College of New York


## Schedule

Week 1: Introduction
Linear, stationary, normal - the stuff biology is not made of.
Week 1-5: Linear systems
Impulse response
Moving Average and Auto Regressive filters
Convolution
Discrete Fourier transform and z-transform
Sampling
Week 6-7: Analog signal processing
Operational amplifier
Analog filtering
Week 8-11: Random variables and stochastic processes
Random variables
Moments and Cumulants
Multivariate distributions, Principal Components
Stochastic processes, linear prediction, AR modeling
Week 12-14: Examples of biomedical signal processing
Harmonic analysis - estimation circadian rhythm and speech
Linear discrimination - detection of evoked responses in EEG/MEGHidden Markov Models and Kalman Filter- identification and filtering

## Jointly Distributed Random Variables (JDRV)

For a family of random variables $X_{1}, \ldots, X_{n}$, which we denote $\boldsymbol{X}$ in short, the joint distribution function is defined as

$$
\begin{aligned}
& F_{X}\left(x_{1}, \ldots, x_{n}\right)=\operatorname{Pr}\left(X_{1} \leqslant x_{1}, \ldots, X_{n} \leqslant x_{n},\right) \\
& F_{X}\left(x_{1}, \ldots,-\infty, \ldots, x_{n}\right)=0, \quad F_{X}(\infty, \ldots, \infty)=1
\end{aligned}
$$

The joint density function is defined as

$$
p_{X}\left(x_{1} \ldots, x_{n}\right)=\frac{\partial^{n} F_{X}\left(x_{1}, \ldots, x_{n}\right)}{\partial x_{1} \ldots \partial x_{n}}
$$

We can "integrate out" or "marginalize" any variable with

$$
p_{X}\left(x_{2}, \ldots, x_{n}\right)=\int_{-\infty}^{\infty} d x_{1} p_{X}\left(x_{1}, x_{2}, \ldots, x_{n}\right)
$$

## JDRV - Conditional distribution

Conditional distribution is the probability distribution of one random variable given that the other has a specific value

$$
F_{X \mid Y}(x \mid y)=\lim _{\epsilon \rightarrow 0} \operatorname{Pr}(X \leqslant x \mid y<Y \leqslant y+\epsilon)
$$

$X \mid Y$ reads " $X$ given $Y$ ". Using the definition of conditional probability

$$
\operatorname{Pr}(A \mid B) \operatorname{Pr}(B)=\operatorname{Pr}(A \wedge B)
$$

We obtain the conditional density of $X$ given $Y$

$$
p_{X \mid Y}(x \mid y)=\frac{p_{X, Y}(x, y)}{p_{Y}(y)}
$$

## JDRV - Conditional distribution

Cartoon example: Dependence of age (x) and hight (y)
"Marginal"
"Joint"




Height distribution of 18 year olds

Height distribution of 5 year olds

Age distribution of entire population

## JDRV - Bayes' Theorem

How do we obtain the conditional distribution of $X \mid Y$ from the conditional distribution of $Y \mid X$ ?

$$
p_{X \mid Y}(x \mid y)=\frac{p_{X, Y}(x, y)}{p_{Y}(y)}=\frac{p_{Y \mid X}(y \mid x) p_{X}(x)}{\int d x p_{X, Y}(x, y)}
$$

Which is another form of the Bayes' Theorem:

$$
p_{X \mid Y}(x \mid y)=\frac{p_{Y \mid X}(y \mid x) p_{X}(x)}{\int d x p_{Y \mid X}(y \mid x) p_{X}(x)}
$$



## JDRV - Independent Random Variables

Random variables are said to be pairwise independent if

$$
p_{X, Y}(x, y)=p_{X}(x) p_{Y}(y)
$$

On equivalently $\quad p_{X \mid Y}(x \mid y)=p_{X}(x)$
Random variables are said to be mutually independent if

$$
p_{X}\left(x_{1,}, \ldots, x_{n}\right)=\prod_{k=1}^{n} p_{X_{k}}\left(x_{k}\right)
$$

Note that mutual independence implies pairwise independence but the inverse is NOT true.


## JDRV - Conditional Independence

Random variables are said to be conditionally independent if

$$
p(x, y \mid z)=p(x \mid z) p(y \mid z)
$$



A sequence is called Markov chain if variables depend only on their immediate predecessor

$$
\begin{gathered}
p\left(x_{1}, \ldots, x_{n}\right)=p\left(x_{n} \mid x_{n-1}\right) p\left(x_{n-1} \mid x_{n-2}\right) \cdots p\left(x_{2} \mid x_{1}\right) p\left(x_{1}\right) \\
X_{1} \cdot X_{2} \rightarrow X_{3} \rightarrow \cdots x_{n 1} \cdot X_{n}
\end{gathered}
$$

The future depends on the past only though the present.

## JDRV - Expectation

Expectation of jointly distributed RVs is defined as

$$
E[f(X, Y)]=\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d x d y f(x, y) p_{X, Y}(x, y)
$$

Consider the sum of random variables $\quad Z=X+Y$
For any random variables we have

$$
E[Z]=E[X]+E[Y]
$$

For independent random variables the PDF of $Z$ is a convolution*

$$
p_{Z}(z)=\int_{-\infty}^{\infty} d x p_{Z \mid X}(z \mid x) p_{X}(x)=\int_{-\infty}^{\infty} d x p_{Y}(z-x) p_{X}(x)
$$

Hence, the sum of independent normal variables remains normal. * Proof using Fourier transforms of pdf (characteristic function) 9

## JDRV - Joint moments

The joint moments of order $n, m$ are defined as

$$
E\left[X^{n} Y^{m}\right]
$$

The most important being the correlation

$$
E[X Y]
$$

For non-zero mean it is better to consider the covariance

$$
\operatorname{cov}[X, Y]=E[(X-E[X])(Y-E[Y])]
$$

To normalize for scale we define the correlation coefficients

$$
\rho_{X, Y}=\frac{\operatorname{cov}[X, Y]}{\operatorname{std}[X] \operatorname{std}[Y]} \quad \text { such that } \quad\left|\rho_{X, Y}\right| \leqslant 1
$$

## JDRV - Independence and correlation

Examples




Rvs are called uncorrelated if $\rho_{X, Y}=0$
Moments of independent RV factor

$$
E\left[X^{n} Y^{m}\right]=E\left[X^{n}\right] E\left[Y^{m}\right]
$$

Independent random variables are uncorrelated. However, uncorrelated random variables are not necessarily independent!

Variances of uncorrelated RVs add:

$$
\operatorname{var}[X+Y]=\operatorname{var}[X]+\operatorname{var}[Y]
$$

## JDRV - Covariance matrix

For a set of RVs, $\boldsymbol{x}=\left[x_{1}, \ldots, x_{n}\right]^{\mathrm{T}}$, we use the covariance matrix*

$$
\boldsymbol{R}_{\boldsymbol{x} x}=E\left[(\boldsymbol{x}-E[\boldsymbol{x}])(\boldsymbol{x}-E[\boldsymbol{x}])^{T}\right]
$$

which reduces for zero mean RVs to

$$
\boldsymbol{R}_{x x}=E\left[\boldsymbol{x} \boldsymbol{x}^{T}\right]
$$

Often the covariance matrix is estimate from a set of $N$ data samples $X=\left[\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{N}\right]$ using the sample averages. For zero-mean variables it is simply:

$$
\hat{\boldsymbol{R}}_{x x}=\frac{1}{N} \sum_{k=1}^{N} \boldsymbol{x}_{k} \boldsymbol{x}_{k}^{T}=\frac{1}{N} \boldsymbol{X} \boldsymbol{X}^{T}
$$

[^0]
## JDRV - Linear regression

Assume the zero-mean RVs $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly related

$$
y=A x
$$

After multiplying with $\boldsymbol{x}^{T}$ and taking the expectation their linear regression coefficients can be found as

$$
\hat{A}=R_{y x} R_{x x}^{-1}
$$

with the equivalent definition for $\boldsymbol{R}_{y x}$ and assuming $\boldsymbol{R}_{x x}$ is invertible. We can then subtract the influence of $\boldsymbol{x}$ on $\boldsymbol{y}$ with

$$
n=y-\hat{A} x
$$

The new variable $\boldsymbol{n}$ is said to be orthogonal to $\boldsymbol{x}$

$$
\operatorname{cov}[\boldsymbol{n}, \boldsymbol{x}]=0
$$

## JDRV - Linear regression

Example: Remove eye blink direction in EEG





## JDRV - Multivariate Normal Distribution

The PDF of normal distributed RVs, $\boldsymbol{x}=\left[x_{1}, \ldots, x_{n}\right]^{\mathrm{T}}$, is given by

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

Its parameters are the mean, $E[\boldsymbol{x}]=\boldsymbol{\mu}$, and covariance, $E\left[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^{T}\right]=\boldsymbol{\Sigma}$
Example:

$$
\begin{aligned}
\mu & =\left[\begin{array}{l}
0 \\
0
\end{array}\right] \\
\Sigma & =\left[\begin{array}{cc}
1 & 0.5 \\
0.5 & 0.4
\end{array}\right]
\end{aligned}
$$




## JDRV - Multivariate Normal Distribution

The marginal PDF of a multivariate normal is also Gaussian. e.g. Marginalizing over $x_{1}$ gives a Gaussian over $\tilde{\boldsymbol{x}}=\left[x_{2}, \ldots, x_{n}\right]^{T}$

$$
p(\boldsymbol{x})=\int_{-\infty}^{\infty} d x_{i} N(\boldsymbol{x}-\mu, \Sigma)=N(\tilde{\boldsymbol{x}}-\tilde{\mu}, \tilde{\Sigma})
$$

with

$$
\tilde{\Sigma}=\left|\begin{array}{cccc}
r_{22} & r_{23} & \cdots & r_{2 n} \\
r_{32} & r_{33} & \cdots & r_{3 n} \\
\vdots & \vdots & \ddots & \vdots \\
r_{n 2} & r_{n 3} & \cdots & r_{n n}
\end{array}\right| \quad \tilde{\mu}=\left|\begin{array}{c}
\mu_{2} \\
\mu_{3} \\
\vdots \\
\mu_{n}
\end{array}\right|
$$

The conditional of a Gaussian, $p(\mathbf{x} \mid \mathbf{y})=p(\mathbf{x}, \mathbf{y}) / \mathrm{p}(\mathbf{y})$, is also Gaussian with the conditional mean computed with linear regression and the covariance given by the Shur complement.

## JDRV - Principal Components

How do we generate samples with given $\boldsymbol{R}_{x x}$ ?
Generate $N$ samples for $n$ independent variables with zero-mean, unit variance: z =randn ( $\mathrm{n}, \mathrm{N}$ )

$$
R_{z z}=I_{n}
$$

We need transformation, $\boldsymbol{x}=\boldsymbol{W} \boldsymbol{z}$, such that

$$
R_{x x}=W R_{z z} W^{T}=W \boldsymbol{W}^{T}
$$

One solution for $\boldsymbol{W}$ is given by the eigenvalue decomposition of $\boldsymbol{R}_{x x}$ with rotation matrix, $\boldsymbol{U}^{-l}=\boldsymbol{U}^{T}$, and diagonal scaling matrix $\boldsymbol{D}$

$$
R_{x x} U=U D \quad W=U D^{1 / 2}
$$

The diagonal elements in D are the eigenvalues of $\boldsymbol{R}_{x x}$, and the columns in $\boldsymbol{U}$ are the principal axes.

## JDRV - Principal Components

Even if the data is not Gaussian it is sometimes useful to consider the transformation

$$
\boldsymbol{z}=\boldsymbol{W}^{-1} \boldsymbol{x}=\boldsymbol{D}^{-1 / 2} \boldsymbol{U}^{T} \boldsymbol{x}
$$

derived from the eigenvalue decomposition of an observed covariance matrix $\boldsymbol{R}_{x x}$.
Columns $\boldsymbol{u}_{k}$ are called the principal components (PC) of samples $\boldsymbol{x}$, and diagonal elements in $\boldsymbol{D}$ are the variance of projections $\boldsymbol{u}_{k}{ }^{T} \boldsymbol{x}$.
Warning: PC not always the right thing to do, e.g. Linear mix of uniform distributed variables.


Assignment 8: Generate figures in slides 14, 15 and 18.

## JDRV - Principal Components Applications

Capture Signal Variance:

- In recordings of data in many dimensions no one single dimension may be very informative.
- Perhaps a combination of directions captures more of the variance in the data.
- Common approach is to looks at projections in the main two principal component directions

$$
\boldsymbol{z}=\left[\boldsymbol{u}_{1} \boldsymbol{u}_{2}\right]^{T} \boldsymbol{x}
$$

## JDRV - Principal Components Applications

Remove Noise Variance

- In some recordings noise is additive to the data.
- Linear combination of electrodes can estimate the noise subspace.
- PCA can be used to find this subspace, say onedimensional in $\mathbf{u}_{1}$,
- The noise can then be filtered spatially by subspace projection

$$
\begin{aligned}
& \boldsymbol{P}=\boldsymbol{I}-\boldsymbol{u}_{1} \boldsymbol{u}_{1}^{T} \\
& \hat{\boldsymbol{x}}=\boldsymbol{P} \boldsymbol{x}
\end{aligned}
$$

Example: 6D Local Field Potentials


To remove subspace $\boldsymbol{V}$ that is not orthonormal:

$$
\begin{equation*}
\boldsymbol{P}=\boldsymbol{I}-\boldsymbol{V}\left(\boldsymbol{V}^{T} \boldsymbol{V}\right)^{-1} \boldsymbol{V}^{T} \tag{20}
\end{equation*}
$$


[^0]:    * Comments on notation: In statistics upper case is used for RV and lower case for a specific value of a RV. Bold face is used to vectors in both cases. In linear algebra bold upper case is used for matrixes and bold lower case for vectors. In signal processing upper case is often reserved for the transform of a time series. When dealing with matrixes I will use the linear algebra convention and the transform domain should be apparent from its argumept 2

