

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 processingHarmonic analysis- estimation circadian rhythm and speechLinear 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 $X_1, ..., X_n$, which we denote X in short, the **joint distribution function** is defined as

$$F_{X}(x_{1},...,x_{n}) = Pr(X_{1} \le x_{1},...,X_{n} \le x_{n},)$$

$$F_{X}(x_{1},...,-\infty,...,x_{n}) = 0, \quad F_{X}(\infty,...,\infty) = 1$$

The joint density function is defined as

$$p_X(x_1,\ldots,x_n) = \frac{\partial^n F_X(x_1,\ldots,x_n)}{\partial x_1\ldots\partial x_n} \qquad \begin{array}{c} X_1 \\ \vdots \\ X_2 \\ \vdots \\ X_3 \end{array}$$

City College of New York

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

$$p_X(x_2,...,x_n) = \int_{-\infty}^{\infty} dx_1 p_X(x_1,x_2,...,x_n)$$
 3



Y

X

JDRV - Conditional distribution

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

$$F_{X|Y}(x|y) = \lim_{\epsilon \to 0} Pr(X \leq x|y < Y \leq y + \epsilon)$$

X|Y reads "X given Y". Using the definition of conditional probability

$$Pr(A|B)Pr(B) = Pr(A \wedge B)$$

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

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



JDRV - Conditional distribution

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

"Marginal"

"Joint"

"Conditional"





JDRV - Bayes' Theorem

How do we obtain the conditional distribution of X|Y from the conditional distribution of Y|X?

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

Which is another form of the **Bayes' Theorem**:

JDRV - Independent Random Variables

Random variables are said to be **pairwise independent** if

City College of New York

On equivalently $p_{X|Y}(x|y) = p_X(x)$

Random variables are said to be **mutually independent** if

$$p_X(x_{1,...,x_n}) = \prod_{k=1}^n p_{X_k}(x_k)$$

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

$$p(x, y|z) = p(x|z) p(y|z)$$



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

$$p(x_1, \dots, x_n) = p(x_n | x_{n-1}) p(x_{n-1} | x_{n-2}) \cdots p(x_2 | x_1) p(x_1)$$

$$(X_1 \to X_2 \to X_3 \to \dots \quad X_{n1} \to X_n)$$

The future depends on the past only though the present.



JDRV - Expectation

Expectation of jointly distributed RVs is defined as

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

Consider the sum of random variables 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*

$$p_{Z}(z) = \int_{-\infty}^{\infty} dx \, p_{Z|X}(z|x) \, p_{X}(x) = \int_{-\infty}^{\infty} dx \, p_{Y}(z-x) \, p_{X}(x)$$

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

 $E[X^nY^m]$

The most important being the **correlation**

E[XY]For non-zero mean it is better to consider the **covariance**

$$cov[X, Y] = E[(X - E[X])(Y - E[Y])]$$

To normalize for scale we define the **correlation coefficients**

$$\rho_{X,Y} = \frac{cov[X,Y]}{std[X]std[Y]}$$

such that $|\rho_{X,Y}| \leq 1$



JDRV - Independence and correlation



Rvs are called **uncorrelated** if $\rho_{X,Y} = 0$

Moments of independent RV factor

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

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

Variances of uncorrelated RVs add:

$$var[X+Y] = var[X] + var[Y]$$
¹¹



JDRV - Covariance matrix

For a set of RVs, $\mathbf{x} = [x_1, ..., x_n]^T$, we use the **covariance matrix**^{*}

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

which reduces for zero mean RVs to

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

Often the covariance matrix is estimate from a set of N data samples $X=[x_1, ..., x_N]$ using the **sample averages**. For zero-mean variables it is simply:

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

* 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 argument?



JDRV - Linear regression

Assume the zero-mean RVs x and y are linearly related

y = A x

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

$$\hat{A} = \boldsymbol{R}_{yx} \boldsymbol{R}_{xx}^{-1}$$

with the equivalent definition for R_{yx} and assuming R_{xx} is invertible. We can then subtract the influence of x on y with

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

The new variable n is said to be orthogonal to x

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



JDRV - Linear regression

Example: Remove eye blink direction in EEG





JDRV - Multivariate Normal Distribution

The PDF of normal distributed RVs, $\boldsymbol{x} = [x_1, ..., x_n]^T$, is given by

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

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



City College of New York

JDRV - Multivariate Normal Distribution

The marginal PDF of a multivariate normal is also Gaussian. e.g. Marginalizing over x_1 gives a Gaussian over $\tilde{x} = [x_2, ..., x_n]^T$

$$p(\mathbf{x}) = \int_{-\infty}^{\infty} dx_i N(\mathbf{x} - \boldsymbol{\mu}, \boldsymbol{\Sigma}) = N(\tilde{\mathbf{x}} - \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$$

with

$$\widetilde{\Sigma} = \begin{bmatrix} r_{22} & r_{23} & \cdots & r_{2n} \\ r_{32} & r_{33} & \cdots & r_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n2} & r_{n3} & \cdots & r_{nn} \end{bmatrix} \qquad \widetilde{\mu} = \begin{bmatrix} \mu_2 \\ \mu_3 \\ \vdots \\ \mu_n \end{bmatrix}$$

The conditional of a Gaussian, $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x},\mathbf{y}) / p(\mathbf{y})$, 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 R_{xx} ? Generate N samples for n independent variables with zero-mean, unit variance: z = randn(n, N)

$$R_{zz} = I_n$$

We need transformation, x = Wz, such that

$$R_{xx} = W R_{zz} W^{T} = W W^{T}$$

One solution for W is given by the eigenvalue decomposition of R_{xx} with rotation matrix, $U^{-1}=U^{T}$, and diagonal scaling matrix D

$$R_{xx}U = UD \qquad W = UD^{1/2}$$

The diagonal elements in D are the eigenvalues of R_{xx} , and the columns in U are the principal axes. 17



JDRV - Principal Components

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

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

derived from the eigenvalue decomposition of an observed covariance matrix R_{xx} .

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

$$\boldsymbol{z} = \left[\boldsymbol{u}_1 \boldsymbol{u}_2 \right]^T \boldsymbol{x}$$

Example: 6D Local Field Potentials



19

City College of New York

JDRV - Principal Components Applications

X6

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 u₁,
- The noise can then be filtered spatially by subspace projection

$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{u}_1 \boldsymbol{u}_1^T$$

 $\hat{x} = P x$

Example: 6D Local Field Potentials



To remove subspace V that is not orthonormal:

$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{V} (\boldsymbol{V}^T \boldsymbol{V})^{-1} \boldsymbol{V}^T$$
 20

