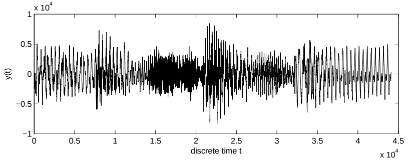
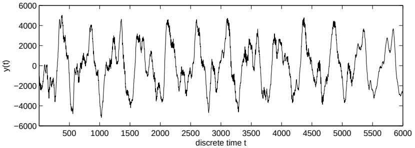
SGN 21006 Advanced Signal Processing: Lecture 2 Random Signals

Ioan Tabus

Department of Signal Processing
Tampere University of Technology
Finland

Studying signal waveforms





(Top) One second of audio signal sampled at 44100 Hz, with 16 bits per sample;

(Bottom) Zoom onto the first 6000 samples

- Finding periodicities
- Finding compact representations:
  - Deterministic function + random component
  - Regression on other signals + random (exploiting correlations to other given signals)
  - A parametric model including random components

Signal representations

- Deterministic + random

$$
\begin{aligned}
X(t) &= f(t) + e(t) \\
&= \sum_{k=1}^{K_1} a_k \sin(\omega_k t + \phi_k) + e_1(t) \qquad \text{Sum of sinusoids} \\
&= \sum_{k=1}^{K_2} b_k \phi_k(t) + e_2(t) \qquad \text{Decomposition on bases } \{\phi_k\}_{k=1}^{K_2}
\end{aligned}
$$

  where $X(t)$ is the given signal, $f(t)$ is a deterministic signal and $e(t), e_1(t), e_2(t)$ are random components ("errors" or "residuals")
- Regression + random

$$
X(t) = \sum_{k=1}^{K_3} X_k(t) + e(t)
$$

  where the random signal $X(t)$ is regressed on other (random) signals $X_1(t), \dots, X_{K_3}(t)$.
- Parametric models
  $X(t) = a_1 X(t-1) + \dots a_{n_a} X(t-n_a) + b_0 e(t) + \dots + b_{n_b} e(t-n_b)$

Random variables

- A random variable X takes values in a set (continuous or discrete). The cumulative distribution function $F_X(x) = Prob(X \leq x)$ can be used for describing probabilities of X falling in an interval: $Prob(a < X \leq b) = F_X(b) - F_X(a)$.
- for a discrete random variable defined on $m, m+1, m+2, \ldots, M$ the probability mass function $p(j) = Prob(X \leq j) - Prob(X < j) = Prob(X = j)$ is easy to use. We have the normalization condition $\sum_{j=m}^{M} Prob(X = j) = 1$.
- continuous random variables $x \in (-\infty, \infty)$ are fully described by the probability density function (pdf) $p(x)$, which obeys the normalization condition $\int_{-\infty}^{\infty} p(x)dx = 1$.
    1. The cumulative distribution function is $F_X(x) = Prob(X \leq x) = \int_{-\infty}^{x} p(y)dy$.
    2. Probability of the variable X falling in the interval $(a, b]$ is $Prob(a < X \leq b) = F_X(b) - F_X(a)$.

Examples of continuous distributions

- Normal distribution, or Gaussian distribution, denoted $\mathcal{N}(\mu, \sigma^2)$, with mean μ and standard deviation σ, with the pdf

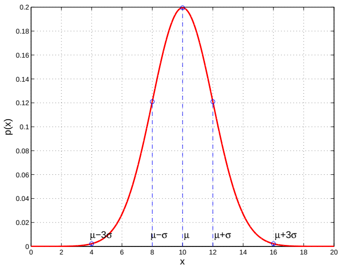
$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

  1. the cumulative distribution

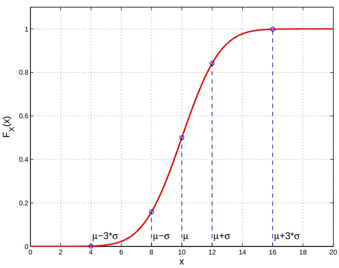
$$F_X(x) = Prob(X \leq x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{\frac{-(y-\mu)^2}{2\sigma^2}} dy$$

     does not have a closed form expression. It is often computed using the cdf of the distribution having parameters $\mu = 0, \sigma = 1$, denoted $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$

Gaussian $\mathcal{N}(\mu = 10, \sigma = 2)$

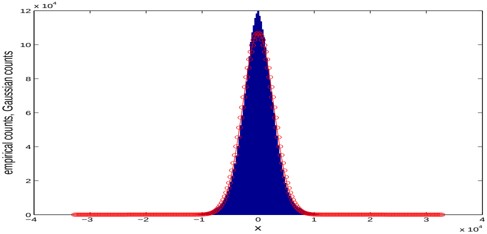


Probability distribution function



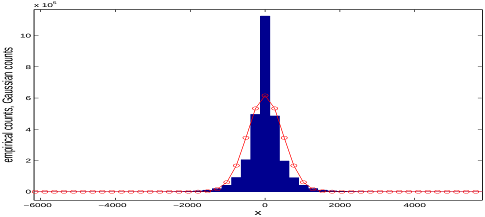
Cumulative distribution function

Fitting a Gaussian distribution $\mathcal{N}(\mu, \sigma)$ to the audio data of Page 1



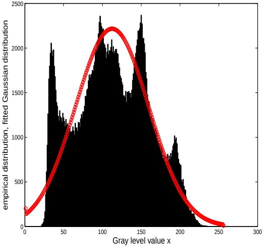
- Estimate the sample mean $\hat{\mu}_1 = \frac{1}{N}\sum_{k=1}^{N} Y(k) = -0.063824 \approx 0$
- Estimate the sample standard deviation $\hat{\sigma}_1 = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(Y(k) - \hat{\mu}_1)^2} = 2742.3$
- Since the range of Y is too large, $Y_k \in \{-2^{15}, 2^{15}\}$, let us take intervals in the range $\{-2^{15}, 2^{15}\}$, of length 256; there are 256 such intervals. Count how many times Y_k falls in a given interval, call it empirical count (blue stems).
- Compute the probability of a Gaussian $\mathcal{N}(\hat{\mu}_1, \hat{\sigma}_1)$ random variable to fall in an interval $(a, b]$, as $Prob(Y \in (a, b]) = F_X(b) - F_X(a)$. Obtain the Gaussian counts as $N \cdot Prob(Y \in (a, b])$, represented in red.
- The empirical counts and the Gaussian counts are reasonably close, so Gaussian distribution is a good approximation. In this representation we have $Y_t = e_t \sim \mathcal{N}(\hat{\mu}_1, \hat{\sigma}_1)$. Is this the best representation?

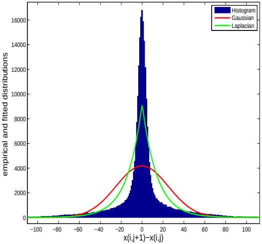
Fitting a Gaussian distribution $\mathcal{N}(\mu, \sigma)$ to the difference $Y(t) - Y(t-1)$



- Estimate the sample mean $\hat{\mu}_2 = \frac{1}{N-1}\sum_{k=2}^{N}(Y(k) - Y(k-1)) = 0.0006464 \approx 0$
- Estimate the sample standard deviation of $e(k) = Y(k) - Y(k-1)$ as $\hat{\sigma}_2 = \sqrt{\frac{1}{N-1}\sum_{k=2}^{N}((Y(k) - Y(k-1)) - \hat{\mu}_2)^2} = 470$
- Repeat the interval construction and the counting as in previous page
- In this representation we have $Y(k) = Y(k-1) + e(k)$ where $e(k) \sim \mathcal{N}(\hat{\mu}_2, \hat{\sigma}_2)$. Here $\hat{\sigma}_2 = 470 << \hat{\sigma}_1 = 2742.3$, so the random component needed for explaining data is much smaller. Prediction by this model will be more accurate!
- The empirical counts and the Gaussian counts are not as close as earlier, so Gaussian distribution may be changed for example to a Laplace distribution.
- Even tighter representations, of the form $Y(t) = a_1 Y(t-1) + \ldots + a_{n_a} Y(t-n_a) + e(t)$ will be discussed in linear prediction lectures.

Fitting a Gaussian distribution $\mathcal{N}(\mu, \sigma)$ to the gray levels on "Barbara" image

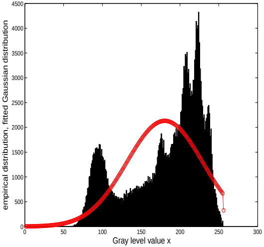


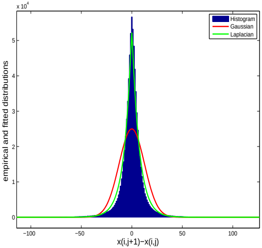


- (Middle) Fitting a Gaussian to the gray levels $y(i,j)$. Estimate the sample mean $\hat{\mu}_2 = \frac{1}{N-1}\sum_{k=2}^{N} Y(k) = 112.45$. Estimate the sample standard deviation $\hat{\sigma} = \sqrt{\frac{1}{N-1}\sum_{k=2}^{N}(Y(k)-\hat{\mu})^2} = 47.2$. Model: $y(i,j) = \varepsilon(i,j) \sim \mathcal{N}(112.4, 47.2)$
- (Right) Fitting Gaussian and Laplace distributions to the differences $y(i,j+1) - y(i,j)$. Interval constructions and counts as in previous page for the Gaussian distribution (red curve). Model: $y(i,j+1) = y(i,j) + \varepsilon(i,j+1)$ with $\varepsilon(i,j+1) \sim \mathcal{N}(-0.01, 25)$.
- Similar estimation of parameters and counting for the Laplace distribution (green curve).

$$p(x;\mu,b) \quad = \quad \frac{1}{2b} e^{-|x-\mu|/b}$$

Fitting a Gaussian distribution $\mathcal{N}(\mu, \sigma)$ to the gray levels on "Lena" image

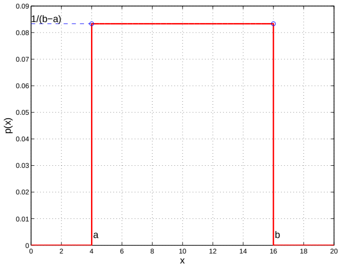


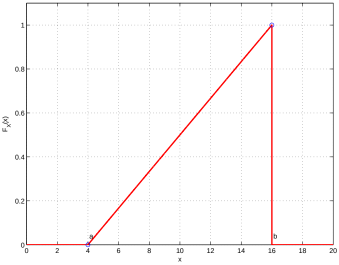


- (Middle) Fitting a Gaussian to the gray levels $y(i,j)$. Estimate the sample mean $\hat{\mu}_2 = \frac{1}{N-1}\sum_{k=2}^{N} Y(k) = 180.2$. Estimate the sample standard deviation $\hat{\sigma} = \sqrt{\frac{1}{N-1}\sum_{k=2}^{N}(Y(k) - \hat{\mu})^2} = 49.05$. Model: $y(i,j) = \varepsilon(i,j) \sim \mathcal{N}(180, 49)$
- (Right) Fitting Gaussian and Laplace distributions to the differences $y(i,j+1) - y(i,j)$. Interval constructions and counts as in previous page for the Gaussian distribution (red curve). Model: $y(i,j+1) = y(i,j) + \varepsilon(i,j+1)$ with $\varepsilon(i,j+1) \sim \mathcal{N}(-0.03, 12.5)$
- Similar estimation of parameters and counting for the Laplace distribution (green curve).

$$p(x;\mu,b) = \frac{1}{2b} e^{-|x-\mu|/b}$$

Uniform distribution $\mathcal{U}(a, b)$





Probability distribution function   Cumulative distribution function

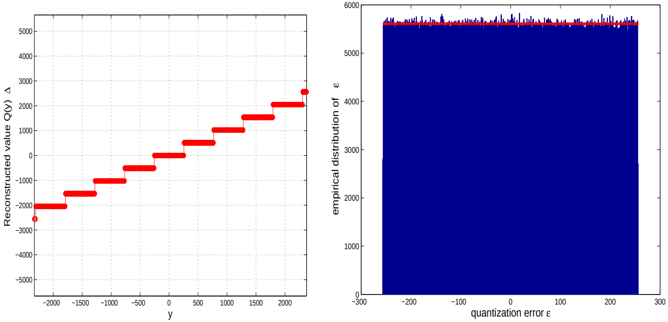
- Uniform distribution, denoted $\mathcal{U}(a, b)$, with mean $\mu = \frac{a+b}{2}$ and standard deviation $\sigma = \sqrt{(b-a)^2/12}$, with the pdf

$$p(x) = \begin{cases} \frac{1}{b-a} & if \quad x \in \{a, b\} \\ 0 & if \quad x \notin \{a, b\} \end{cases}$$

- the cumulative distribution

$$F_X(x) = \begin{cases} \frac{x-a}{b-a} & if \quad x \in \{a, b\} \\ 0 & if \quad x \notin \{a, b\} \end{cases}$$

Uniform quantization



- The quantized value is computed as $Q(y) = \text{round}\ (y/\Delta)$ and has a dynamic range Δ times smaller than the initial dynamic range.
- The reconstructed value is $\hat{y} = Q(y)\Delta$. The quantization error is $\varepsilon = y - \hat{y}$.
- For the audio signal we have $y \in \{-2^{15}, 2^{15}\}$. Let's take $\Delta = 512$.
- The quantization error ε belongs to $\{-256, \ldots, 256\}$, hence the uniform distribution $\mathcal{U}(a, b)$ has parameters $a = -256, b = 256$. The histogram of ε is shown in blue, while the ideal uniform counts, corresponding to the uniform distribution $\mathcal{U}(a, b)$, are shown in red.

Joint probabilities; Vector random variables

- the joint probability of X and Y is $p(x,y) = Prob(X = x; Y = y)$
- Two situations:
  1. X and Y are independent iff for all x, y
     $$Prob(X = x; Y = y) \quad = \quad Prob(X = x)Prob(Y = y)$$
  2. when X and Y are NOT independent, the factorization involves conditional probabilities, by the rule of Bayes
     $$\begin{aligned} Prob(X = x; Y = y) &= Prob(X = x|Y = y)Prob(Y = y) \\ &= Prob(Y = y|X = x)Prob(X = x) \end{aligned}$$
- for a vector of random variables, $\underline{X} = [X_1, X_2, \ldots, X_n]^T$, having pdf $p(\underline{x})$, the mean is
  $$\underline{\mu} = E[\underline{X}] = \int \underline{x}p(\underline{x})dx_1 \ldots dx_n,$$
  the correlation matrix is
  $$R \quad = \quad E[\underline{x}\underline{x}^T]$$
  and the covariance matrix is
  $$\Sigma \quad = \quad E[(\underline{x} - E[\underline{x}])(\underline{x} - E[\underline{x}])^T = R - \underline{\mu}\underline{\mu}^T$$

Gaussian random vectors

- consider the vector of random variables, $\underline{X} = [X_1, X_2, \ldots, X_n]^T$, having the Gaussian pdf

$$p(\underline{x}) = \frac{1}{(2\pi)^{N/2}(\det\Sigma)^{1/2}} e^{-(\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu})/2}$$

  with mean $\underline{\mu}$ and covariance matrix Σ.

- A linear transformation B applied to a Gaussian vector $\underline{x}$ of mean $\underline{\mu}_x$ and covariance matrix Σ_x, results in a Gaussian vector $\underline{y} = B\underline{x}$, having the mean

$$\underline{\mu}_y = B\underline{\mu}_x$$

  and covariance matrix

$$\Sigma_y = B\Sigma_x B^T$$

Uncorrelated Gaussian random vectors

- If the n_x-vector $\underline{x}$ and n_y-vector $\underline{y}$ are zero mean and uncorrelated and jointly Gaussian, then they are also independent.
  Proof: Denote the joint vector $\underline{z} = [\underline{x}^T \quad \underline{y}^T]^T$. The Gaussian assumption states:

$$\begin{aligned} p(\underline{x}) &= \frac{1}{(2\pi)^{n_x/2}(\det \Sigma_x)^{1/2}} e^{-\underline{x}^T \Sigma_x^{-1} \underline{x}/2} \\ p(\underline{y}) &= \frac{1}{(2\pi)^{n_y/2}(\det \Sigma_y)^{1/2}} e^{-\underline{y}^T \Sigma_y^{-1} \underline{y}/2} \\ p(\underline{z}) &= \frac{1}{(2\pi)^{(n_x+n_y)/2}(\det \Sigma_z)^{1/2}} e^{-\underline{z}^T \Sigma_z^{-1} \underline{z}/2} \end{aligned}$$

We need to show that $p(\underline{z}) = p(\underline{x})p(\underline{y})$.
If $\underline{x}$ and $\underline{y}$ are uncorrelated $E[\underline{x}\underline{y}^T] = 0$. The covariance matrix of $\underline{z}$ is

$$\Sigma_z = E[\begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix} \begin{bmatrix} \underline{x}^T & \underline{y}^T \end{bmatrix}] = \begin{bmatrix} E[\underline{x}\underline{x}^T] & E[\underline{x}\underline{y}^T] \\ E[\underline{y}\underline{x}^T] & E[\underline{y}\underline{y}^T] \end{bmatrix} = \begin{bmatrix} E[\underline{x}\underline{x}^T] & 0 \\ 0 & E[\underline{y}\underline{y}^T] \end{bmatrix} = \begin{bmatrix} \Sigma_x & 0 \\ 0 & \Sigma_y \end{bmatrix}$$

The quadratic form in the exponential of $p(\underline{z})$ is

$$\underline{z}^T \Sigma_z^{-1} \underline{z} = \begin{bmatrix} \underline{x}^T & \underline{y}^T \end{bmatrix} \begin{bmatrix} \Sigma_x & 0 \\ 0 & \Sigma_y \end{bmatrix}^{-1} \begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix} = x^T \Sigma_x^{-1} \underline{x} + y^T \Sigma_y^{-1} \underline{y}$$

and hence

$$e^{-z^T \Sigma_z^{-1} \underline{z}/2} = e^{-(x^T \Sigma_x^{-1} \underline{x} + y^T \Sigma_y^{-1} \underline{y})/2} = e^{-x^T \Sigma_x^{-1} \underline{x}/2} e^{-y^T \Sigma_y^{-1} \underline{y}/2}$$

The determinant of Σ_z also factorizes as $\det \Sigma_z = \det \Sigma_x \det \Sigma_y$ and finally $p(\underline{z}) = p(\underline{x})p(\underline{y})$ which means $\underline{x}$ and $\underline{y}$ are independent.

Expectation for continuous variables

- for continuous random variables $x \in (-\infty, \infty)$, fully described by the probability density function (pdf) $p(x)$ the expectation of a function $g(X)$ is $E[g(X)] = \int_{-\infty}^{\infty} g(x)p(x)dx$
- Important expectations
  1. mean, or first moment, or expected value of X
  $$\mu = E[X] = \int_{-\infty}^{\infty} xp(x)dx$$
  2. variance, or expected value of $(X-\mu)^2$
  $$\sigma^2 = var(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx$$
  3. second moment, or expected value of X^2
  $$E[X^2] = \int_{-\infty}^{\infty} x^2 p(x)dx = \sigma^2 + \mu^2$$

Expectation for discrete variables

- for discrete random variables X taking values in the set $m, m+1, m+2, \ldots, M$ with the probability mass function $p(j) = Prob(X \leq j) - Prob(X < j) = Prob(X = j)$ the expectation of a function $g(X)$ is $E[g(X)] = \sum_{j=m}^{M} g(j)p(j)$
- Important expectations
  1. mean, or first moment, or expected value of X
  $$\mu = E[X] = \sum_{j=m}^{M} jp(j)$$
  2. variance, or expected value of $(X - \mu)^2$
  $$\sigma^2 = var(X) = E[(X - m)^2] = \sum_{j=m}^{M} (j - \mu)^2 p(j)$$
  3. second moment, or expected value of X^2
  $$E[X^2] = \sum_{j=m}^{M} j^2 p(j) = \sigma^2 + \mu^2$$

Properties of expectation operator

For simplicity we take the case of discrete random variables, but the results are holding for continuous random variables as well.

- The linearity property: given two random variables, X and Y, with joint pmf $g(x,y) = Prob(X = x; Y = y)$ and two constants a and b, then

  $$\boxed{\mathrm{E[aX+bY] = aE[X] + bE[Y]}}$$

  Proof:

  $$\begin{aligned} E[aX+bY] &= \sum_x \sum_y (ax+by) Prob(X=x; Y=y) \\ &= a\sum_x \sum_y x Prob(X=x; Y=y) + b\sum_x \sum_y y Prob(X=x; Y=y) \\ &= a\sum_x x Prob(X=x) + b\sum_y y Prob(Y=y) = aE[X] + bE[Y] \end{aligned}$$

- The expectation of a product of two independent random variables is equal to the product of expectations $E[XY] = E[X]E[Y]$
  Proof: If X is independent of Y, then
  $Prob(X = x, Y = y) = Prob(X = x)Prob(Y = y)$ and

  $$\begin{aligned} E[XY] &= \sum_x \sum_y xy Prob(X=x; Y=y) \\ &= \sum_x x Prob(X=x) \sum_y y Prob(Y=y) = E[X]E[Y] \end{aligned}$$

Properties of expectation operator

- In general the expectation of a product of two random variables is NOT equal to the product of expectations, $E[XY] \neq E[X]E[Y]$ unless the random variables are independent.
- The difference $E[XY] - E[X]E[Y]$ is equal to the crosscorrelation function

$$\begin{aligned} R(X,Y) &= E[(X - E[X])(Y - E[Y])] = E[XY - YE[X] - XE[Y] + E[X]E[Y]] \\ &= E[XY] - 2E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y] \end{aligned}$$

- For independent variables the cross-correlation is 0.
  Proof: $R(X,Y) = E[XY] - E[X]E[Y] = 0$.

Gaussian vectors with diagonal covariance matrix have independent components

- Take a random vector $\underline{X} = [X_1 \quad \dots \quad X_n]^T$ with mean $\underline{\mu} = [\mu_1 \quad \dots \quad \mu_n]^T$ and diagonal covariance matrix

$$R = E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})^T] = \begin{bmatrix} E[(X_1 - \mu_1)^2] & 0 & \dots & 0 \\ 0 & E[(X_2 - \mu_2)^2] & \dots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & E[(X_n - \mu_n)^2] \end{bmatrix}$$

Proof that Gaussian vectors with diagonal covariance matrix have independent components:

$$R^{-1} = \left(E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})^T]\right)^{-1} = \begin{bmatrix} 1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & \dots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1/\sigma_n^2 \end{bmatrix}$$

Hence $(\underline{x} - \underline{\mu})^T R^{-1} (\underline{x} - \underline{\mu}) = \sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma_i^2}$ and $\det R = \prod_{i=1}^n \sigma_i^2$

$$\begin{aligned} p(\underline{x}) &= \frac{1}{(2\pi)^{n/2}(\det R)^{1/2}} e^{-(\underline{x}-\underline{\mu})^T R^{-1}(\underline{x}-\underline{\mu})/2} = \frac{1}{(2\pi)^{n/2}(\det R)^{1/2}} e^{-\sum_{i=1}^n \frac{(x_i-\mu_i)^2}{2\sigma_i^2}} \\ &= \prod_{i=1}^n \frac{1}{(2\pi)^{1/2}\sigma_i} e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}} = \prod_{i=1}^n p(x_i) = p(x_1, x_2, \dots, x_n) \rightarrow \text{INDEPENDENCE} \end{aligned}$$

Uncorrelated jointly Gaussian variables are independent

- Take $X_1 \quad \ldots \quad X_n$ uncorrelated jointly Gaussian.
- "uncorrelatedness" means that $E[(X_i - \mu_i)(X_j - \mu_j)] = 0$ for all i, j, thus the covariance matrix of the vector $[X_1 \; X_2 \; \ldots X_n]$ is diagonal.
- "jointly Gaussian" means that $p(\underline{x}) = \frac{1}{(2\pi)^{n/2}(\det R)^{1/2}} e^{-(\underline{X}-\underline{\mu})^T R^{-1}(\underline{X}-\underline{\mu})/2}$.
- From the previous page it results that $X_1 \quad \ldots \quad X_n$ are independent.
- Hence uncorrelated Gaussian variables are independent
- The opposite is true in general: independent variables are uncorrelated
- To the extent that the distributions can be approximated well by Gaussiaan distribution, one can identify in general "uncorrelated" and "independent", at least at the level of heuristic descriptions.

De-correlating random vectors: the Karhunen-Loeve transform

- Consider a random vector $\underline{X} = [X_1 \quad \dots \quad X_n]^T$ with mean $\underline{\mu} = [\mu_1 \quad \dots \quad \mu_n]^T$ with an arbitrary covariance matrix R_X, which is not a diagonal matrix.
- The non-diagonal element (i,j) of the matrix R_X is

$$R_X(i,j) = E[(X_i - \mu_i)(X_j - \mu_j)^T]$$

- The random variables are not independent! We want to find, by using a simple linear transformation, a vector of n random variables which are uncorrelated (and if Gaussian distributed, are also independent).
- We want to find the transformation matrix B so that $\underline{Y} = B(\underline{X} - \mu)$ has the elements uncorrelated with each other. That will happen if the covariance matrix $E[\underline{Y}\underline{Y}^T]$ is diagonal.

$$\begin{aligned} R_Y &= E[\underline{Y}\underline{Y}^T] = E[B(\underline{X} - \mu)(B(\underline{X} - \mu))^T] \\ &= E[B(\underline{X} - \mu)(\underline{X} - \mu)^T B^T] = E[BR_X B^T] \end{aligned}$$

In general the matrix B will be complex valued, so all transposition operations $(\cdot)^T$ must be replaced by "complex conjugation and transpose" operation $(\cdot)^H$.

$$R_Y = E[\underline{Y}\underline{Y}^H] = E[BR_X B^H]$$

The Karhunen-Loeve transform

- Consider the eigenvalue decomposition of R_x, where each eigenvalue-eigenvector pair $\lambda_i, \underline{q}_i$ obeys $R_x \underline{q}_i = \lambda_i \underline{q}_i$. Denote $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$. One can choose such eigenvectors that the matrix $Q = [\underline{q}_1 \; \underline{q}_2 \; \ldots \underline{q}_n]$ is unitary, so that

$$\begin{aligned} Q^H Q &= I \\ R_x Q &= Q\Lambda \\ R_x &= Q\Lambda Q^H \\ \Lambda &= Q^H R_x Q \end{aligned}$$

- Taking $B = Q^H$ we have the desired transformation

$$\begin{aligned} \underline{Y} &= Q^H(\underline{X} - \underline{\mu}) \\ R_Y &= E[\underline{Y}\,\underline{Y}^H] = Q^H R_X Q = \Lambda \end{aligned}$$

  and, since the matrix Λ is diagonal, the components of the vector $\underline{Y}$ are not correlated, $E[Y_i Y_j] = 0$.

- if the distribution of the initial vector $\underline{X}$ was Gaussian, also the transformed vector $\underline{Y}$ is Gaussian distributed, and hence the components of the vector $\underline{Y}$ are independent!
- the dependent components of $\underline{X}$ are transformed easily by the KL transform into independent components, over which the study is much simpler, since each component can be studied separately!
- the results of the analysis over the components of $\underline{Y}$ can then be phrased in terms of the initial random variables in $\underline{X}$ by the inverse transform

$$\underline{X} = Q\underline{Y} + \mu$$

Simulation of random variables in Matlab

- Uniform distribution $\mathcal{U}(0,1)$ on the open interval (0,1):
  rand(n) returns a n-by-n matrix containing pseudorandom values drawn from the standard uniform distribution on the open interval (0,1). rand(m,n) or rand([m,n]) returns an m-by-n matrix.
- normal distribution $\mathcal{N}(0,1)$ of zero mean and standard deviation $\sigma = 1$:
  r = randn(n) returns an n-by-n matrix containing pseudorandom values drawn from the standard normal distribution. randn(m,n) or randn([m,n]) returns an m-by-n matrix.
- Inverse transform method
  We want to generate n samples from any given cumulative distribution function, say $F_X(x)$. The inverse method requires to generate n samples $u_1, \dots, u_n$ from the uniform distribution $\mathcal{U}(0,1)$ (e.g. by using *rand(n)* function) and then obtain the desired samples as $x_1 = F_X^{-1}(u_1), \dots, x_n = F_X^{-1}(u_n)$, where the function F_X^{-1} is defined as

$$F_X^{-1}(u) = \arg\min_x \{F(x) \geq u, u \in [0,1]\}$$

Proof: We need to show that $Prob(X \leq x) = F_X(x)$, where $X = F_X^{-1}(U)$ and where U satisfies $Prob(U \leq u) = u$. We evaluate: $Prob(X \leq x) = Prob(F_X^{-1}(U) \leq x)$. If the function $F_X(x)$ is continuous, the function $F_X^{-1}(\cdot)$ is the proper inverse function of $F_X(x)$.
$Prob(X \leq x) = Prob(F_X^{-1}(U) \leq x) = Prob(U \leq F_X(x)) = F_X(x)$. The case of discontinuous $F_X(x)$ can be shown to lead to same result, that $F_X^{-1}(U)$ is distributed as $F_X(\cdot)$.