ECE 6960: Adv. Random Processes & Applications
Lecture Notes, Fall 2010

Lecture 6

Today: (1) Gaussian R.V.s, (2) Linear transforms of R.V.s

- HW 2 due today; HW 3 assigned and due Sept 16, 5pm.

1 Joint Gaussian R.V.s

<u>Def'n:</u> *Joint Gaussian R.V.s.*
An n-length R.V. $\mathbf{X}$ is jointly Gaussian with mean $\mu_{\mathbf{X}}$, and covariance matrix $C_{\mathbf{X}}$ if it has the pdf,

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n det(C_{\mathbf{X}})}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{X}})^T C_{\mathbf{X}}^{-1}(\mathbf{x}-\mu_{\mathbf{X}})\right]$$

where $det()$ is the determinant of the covariance matrix, and $C_{\mathbf{X}}^{-1}$ is the inverse of the covariance matrix.

Note: The 'Inner Product' means that the transpose is in the middle. You will get one number out of an inner product. The 'Outer Product' means the transpose is on the outside of the product, and you will get a matrix out of it.

Example: Find the pdf of $n = 2$ Joint Gaussian r.v.s:

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi\tilde{\sigma}_2^2}} e^{-\frac{(x_2-\tilde{\mu}_2(x_1))^2}{2\tilde{\sigma}_2^2}}$$

where

$$\tilde{\mu}_2(x_1) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \quad \tilde{\sigma}_2^2 = \sigma_2^2(1-\rho^2)$$

Properties:

- Any marginal pdf of a joint Gaussian R.V. must also be (multivariate) Gaussian.
- The conditional pdf of a joint Gaussian R.V. is also Gaussian.
- A linear combination of a joint Gaussian R.V. is also joint Gaussian. This includes the sum of the elements of the joint Gaussian R.V.

Once you know that a R.V. is jointly Gaussian, all you need to do is find its mean vector and covariance matrix to completely specify its distribution. Also useful to disprove that a R.V. is jointly Gaussian. The alternate definition in Leon-Garcia is:

Def'n: *Multivariate Gaussian*
$\mathbf{X}$, an n-length R.V., is a jointly Gaussian random vector if and only if every linear combination $Z = \mathbf{a}^T\mathbf{X}$, for $\mathbf{a} \in \Re^n$, is a Gaussian random variable.

Example: Not Jointly Gaussian
: Let X be unit-variance, zero-mean, Gaussian r.v. Let Y be defined as:

$$Y = \begin{cases} -X, & \text{if } |X| < c \\ X, & \text{if } |X| \geq c \end{cases}$$

where $c > 0$. Note Y also has the Gaussian distribution, so both X and Y are Gaussian. What happens to the linear combination $Z = X + Y$? When $|X| < c$, $Y = -X$ so $Z = X + Y = 0$. When $|X| > c$, $Y = X$ so $Z = X + Y = X + X = 2X$. So the distribution of the linear combination is a mixture, with positive probability of $Z = 0$, clearly not a Gaussian distribution. Thus the vector $[X, Y]^T$ is not a multivariate Gaussian R.V. Note that you can even set c such that the correlation of X and Y is zero [3]. So one cannot say that uncorrelated Gaussian r.v.s are in fact independent!

2 Linear Transforms of R.V.s

Recall the form of the general linear transform, as described in lecture 5:

$$\mathbf{Y} = A\mathbf{X} \tag{1}$$

where A is a $n \times n$ matrix of constants. In this section, we discuss the mean vector and covariance matrix of $\mathbf{Y}$. As with constants multiplying a random variable, a matrix of constants can be removed from the expected value. So

$$\mu_{\mathbf{Y}} = E_{\mathbf{Y}}\left[\mathbf{Y}\right] = E_{\mathbf{X}}\left[A\mathbf{X}\right] = AE_{\mathbf{X}}\left[\mathbf{X}\right] = A\mu_{\mathbf{X}} \tag{2}$$

To find the covariance matrix, start from the definition and plug in for $\mathbf{Y}$ and $\mu_{\mathbf{Y}}$,

$$\begin{aligned} \text{Cov}\left(\mathbf{Y}\right) &= E_{\mathbf{Y}}\left[(\mathbf{Y} - \mu_{\mathbf{Y}})(\mathbf{Y} - \mu_{\mathbf{Y}})^T\right] \\ &= E_{\mathbf{X}}\left[(A\mathbf{X} - A\mu_{\mathbf{X}})(A\mathbf{X} - A\mu_{\mathbf{X}})^T\right] \end{aligned}$$

We can "take out" the A from each expression, but be sure to pull it out from the side it is "facing". Also, recall $(CD)^T = D^TC^T$, that is, you must reverse the order of the factors when taking the transpose of a product. So,

$$\begin{aligned} \mathrm{Cov}\left(\mathbf{Y}\right) &= E_{\mathbf{X}}\left[A(\mathbf{X}-\mu_{\mathbf{X}})(\mathbf{X}-\mu_{\mathbf{X}})^T A^T\right] \\ &= AE_{\mathbf{X}}\left[(\mathbf{X}-\mu_{\mathbf{X}})(\mathbf{X}-\mu_{\mathbf{X}})^T\right] A^T \\ &= A\mathrm{Cov}\left(\mathbf{X}\right) A^T \end{aligned} \tag{3}$$

We can use (2) and (3) to our benefit.

1. We can transform independent r.v.s to a vector with an arbitrary covariance matrix and mean vector.

2. We can transform any R.V.s with known mean and covariance matrix to be a vector of independent r.v.s.

2.1 Generating R.V.s with Arbitrary Covariance

The first is useful, for example, in Matlab for the generation of correlated random vectors. Assume that $\mu_{\mathbf{X}} = \mathbf{0}$ and $\mathrm{Cov}\left(\mathbf{X}\right) = I$, that is, $\mathbf{X}$ is composed of i.i.d. unit variance zero mean random variables. Matlab by default generates i.i.d. random variables, either Gaussian or uniform. Then $\mathbf{Y}$ is also zero mean and from (3, $\mathrm{Cov}\left(\mathbf{Y}\right) = AA^T$. So, all we need to do to achieve an arbitrary covariance matrix $\mathrm{Cov}\left(\mathbf{Y}\right)$ is to find a matrix A such that $\mathrm{Cov}\left(\mathbf{Y}\right) = AA^T$. As it turns out, valid covariance matrices are symmetric positive semidefinite matrices, and thus have a square root. So if we write $C_{\mathbf{Y}} = \mathrm{Cov}\left(\mathbf{Y}\right)$, we can write its square root as $C_{\mathbf{Y}}^{1/2}$. We'd then assign $A = C_{\mathbf{Y}}^{1/2}$ to get $\mathbf{Y}$ to have this covariance matrix.

Let's consider the second application before finishing this thought.

2.2 Decorrelation of R.V.s

Here, we discuss the setting of A in the transform $\mathbf{Y} = A\mathbf{X}$ to achieve a diagonal covariance matrix for $\mathbf{Y}$, which indicates that all pairs of components (i, j) with $i \neq j$ have zero covariance. In other words, all pairs of r.v.s in the R.V. are uncorrelated. In short, we say $\mathbf{Y}$ is an uncorrelated R.V. Why should we want this? Let's go back to those examples.

- Multiple antenna transceivers. The channels between each pair of antennas cause one linear transform H. Assuming that the transmit antennas sent independent r.v.s at each time, we might determine H^{-1} by finding the matrix which decorrelates the R.V.s measured at the receive antenna.

- Secret key generation from fading measurements. We can eliminate correlation between channel samples (RSS measurements) over time to generate independent samples; this makes a resulting secret key stronger.

- Finance. Come up with mutual funds which are uncorrelated; thus allowing explicit *diversification*. See lecture 14 of ECE 5510, Fall 2009.
- Finite impulse response (FIR) filters. Come up with what is called a "whitening filter", which takes correlated noise and spreads it across the frequency spectrum.

Let's review the goal, which is to find a matrix A for a linear transformation $\mathbf{Y} = A\mathbf{X}$ which causes $C_{\mathbf{Y}}$ to be a diagonal matrix.

2.2.1 Singular Value Decomposition (SVD)

The solution is to use the singular value decomposition (SVD). It says that any matrix C can be written as

$$C = U\Lambda V^T$$

where U and V are unitary matrices, and Λ is a diagonal matrix. (A unitary matrix is one that has the property that $U^T U = I$, the identity matrix. It is an orthogonal transformation, if you've heard of that.)

Covariance matrices like $C_{\mathbf{X}}$ have two properties which make things simpler for us: it is symmetric and positive semi-definite. This simplifies the result; it means that $V = U$ in the above equation, so

$$C_{\mathbf{X}} = U\Lambda U^T$$

Also, for positive semi-definite matrices, all of the diagonal elements of Λ are non-negative. The columns of U are called the eigenvectors and the diagonal elements of Λ are called the eigenvalues. Specifically, $\Lambda = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}$. Using this linear algebra result, we can come up with the desired transform. The answer is as follows:

1. Take the SVD of the known covariance matrix $C_{\mathbf{X}}$ to find U and Λ.
2. Let $A = U^T$, that is, define vector $\mathbf{Y}$ as $\mathbf{Y} = U^T\mathbf{X}$.

What happens then? Well, we know that

$$C_{\mathbf{Y}} = AC_{\mathbf{X}}A^T = U^T C_{\mathbf{X}} U = U^T U\Lambda U^T U$$

since we can re-write $C_{\mathbf{X}}$ as $U\Lambda U^T$. Next, Since U is a unitary matrix, $C_{\mathbf{Y}} = I\Lambda I = \Lambda$. We now have a transformed R.V. $\mathbf{Y}$ with a diagonal covariance matrix, in other words, an uncorrelated random vector!

What is the square root of $C_{\mathbf{X}}$? First, Λ has a square root: since it is diagonal, with elements λ_i, I can form $\Lambda^{1/2}$ with diagonal elements $\lambda_i^{1/2}$, which would then achieve $\Lambda = (\Lambda^{1/2})(\Lambda^{1/2})^T$, our definition of a square root matrix. Then

$$C_{\mathbf{X}} = U\Lambda U^T = U(\Lambda^{1/2})(\Lambda^{1/2})^T U^T = (U\Lambda^{1/2})(U\Lambda^{1/2})^T = (C_{\mathbf{X}}^{1/2})(C_{\mathbf{X}}^{1/2})^T \quad (4)$$

where we now have $C_{\mathbf{X}}^{1/2} = U\Lambda^{1/2}$.

Note that Matlab calculates the SVD with command

```
[U, S, V] = svd(C_X);
```

HW Problem: Find A in $\mathbf{y} = A\mathbf{X}$ that makes $C_{\mathbf{Y}} = I$, given arbitrary covariance matrix $C_{\mathbf{X}} = \mathrm{Cov}\,(\mathbf{X})$, and prove that your choice makes $C_{\mathbf{Y}} = I$.

Example: Average and Difference of Two i.i.d. R.V.s

We represent the arrival of two people for a meeting as r.v.s X_1 and X_2. Let $\mathbf{X} = [X_1, X_2]^T$. Assume that the two people arrive independently, with the same variance σ^2 and mean μ. Consider the average arrival time $Y_1 = (X_1 + X_2)/2$, and the difference between the arrival times, $Y_2 = X_1 - X_2$. The latter is a wait time that one person must wait before the second person arrives. Show that the average time and the difference between the times are uncorrelated.

You can take these steps to solve this problem:

1. Let $\mathbf{Y} = [Y_1, Y_2]^T$. What is the transform matrix A in the relation $\mathbf{Y} = A\mathbf{X}$?
2. What is the mean matrix $\mu_{\mathbf{Y}} = E_{\mathbf{Y}}\,[\mathbf{Y}]$? (Note this isn't really needed to answer the question, but is good practice anyway.)
3. What is the covariance matrix of $\mathbf{Y}$?
4. How does the covariance matrix show that the two are uncorrelated?

2.3 Compression using the Karhunen-Loeve Transform (KLT)

The multiplication by U^T is also called the Karhunen-Loeve Transform (KLT). The idea is that the column vectors of U, let's call them $\mathbf{u}_i$,

$$U = [\mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_n]$$

form an orthogonal basis for the vector $\mathbf{X}$. We can transform it into uncorrelated components by multiplication by U^T, and then transform it back to its original values by multiplying again by U. But some of the uncorrelated components vary more than others. The variance of the components of $\mathbf{Y} = [y_1, \ldots, y_n]^T$ are given by

$\lambda_1, \ldots, \lambda_n$, respectively. So, if we need a $n-k$-dimensional representation of the original vector $\mathbf{X}$, we would need to ignore k components, and it would be best to ignore the components which have the least variance.

To do this we would:

1. Measure the n-dimensional vector $\mathbf{X}$.

2. Compute and store (or send) the $n-k$-dimensional vector $\hat{\mathbf{Y}}$ as
$$\hat{\mathbf{Y}} = P^T\mathbf{X} \quad \text{where } P = [\mathbf{u}_1 \mid \cdots \mid \mathbf{u}_{n-k}].$$
Note P is a $n \times n-k$ matrix.

3. When required to recreate the original vector, we'd provide our best approximation as:
$$\hat{\mathbf{Y}} = P\hat{\mathbf{Y}}.$$

What we've done is to zero out orthogonal components which have very low variance. The approximation $\hat{\mathbf{X}}$ is the "best" $n-k$ dimensional representation, in the mean-squared error sense, to the original $\mathbf{X}$ [1].

The KLT has application in compression for noise reduction, spectral estimation [4], storage and communication, and anomaly detection [2].

References

[1] B. N. Datta. *Applied and computational control, signals, and circuits, Volume 1.* Birkhuser, 1999.

[2] A. Lakhina, M. Crovella, and C. Diot. Diagnosing network-wide traffic anomalies. In *ACM SIGCOMM*, Aug. 2004.

[3] E. Melnick and A. Tenenbein. Misspecifications of the normal distribution. *American Statistician*, pages 372–373, 1982.

[4] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986.