ESE 520 Probability and Stochastic Processes

Lecture 10

"Covariance matrix. Bivariate normal distribution"

Now we extend the notion of the covariance to a random vector of any dimension $n, n = 1, 2, ...$

Definition 1. Let $X = (X_1, X_2, ..., X_n)$ be a random vector defined on a probability space $(\Omega, \mathcal{F}, P)$. The matrix $R = (cov(X_i, X_j)), i, j = 1, 2, ..., n$ is then called the *covariance matrix* of X.

Otherwise expressed:

$$R = E[(X - \mu)(X - \mu)^T]$$

where, again, any vector $x \in \mathbb{R}^n$ is understood here as a column vector (a matrix of size $n \times 1$) so that x^T is a row vector (a matrix of size $1 \times n$).

It follows from the definition that:

i) a covariance matrix R is symmetric;

ii) R is a *positive semidefinite matrix*; that is, for any $x \in \mathbb{R}^n$:

$$x^T R x = E[x^T (X - \mu)(X - \mu)^T x] = E[\|(X - \mu)x\|^2] \geq 0,$$

where $\|y\|$ denotes the Euclidean norm of a vector y.

Proposition 1. *Let $X = (X_1, X_2, ..., X_n)$ be a non-degenerate Gaussian random vector with the density*

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det R}} \exp\{-\frac{1}{2}(x - \mu)^T R^{-1}(x - \mu)\}.$$

Then $E(X) = \mu$ and R is the covariance matrix of X.

Proof. First we notice that since R is a positive definite matrix there exists a positive definite "square root matrix of R" which we will denote by $R^{1/2}$.

Indeed, any positive definite matrix can be diagonalized in the form

$$R = S\Lambda S^T,$$

where $\Lambda = (\lambda_{ij})$ is a diagonal matrix with the entries λ_{ii} being equal to the eigenvalues of R (all strictly positive numbers) and S is an invertible matrix satisfying the property $S^{-1} = S^T$. Because of that,

$$R^{1/2} := S\Lambda^{1/2}S^T, \tag{1}$$

where $\Lambda^{1/2}$ is a diagonal matrix with the entries $\sqrt{\lambda_{ij}}$.

We also note that by construction $R = R^{1/2}R^{1/2}$.

For compactness, we do calculations in a matrix form:

$$E[(X-\mu)(X-\mu)^T] =$$

$$\int_{\mathbb{R}^n} \frac{(x-\mu)(x-\mu)^T}{\sqrt{(2\pi)^n \det R}} \exp\{-\frac{1}{2}(x-\mu)^T R^{-1}(x-\mu)\}dx \tilde{=}$$

Now: we make change of variables

$$z := R^{-1/2}(x-\mu) \tag{2}$$

so that

$$dz = (\det R^{-1/2})dx,$$

where $(\det R^{-1/2})$ is the Jacobian of the linear transformation in (2). Using (1), one can see that $(\det R^{-1/2}) = \frac{1}{\sqrt{\det R}}$ so that

$$dz = \frac{1}{\sqrt{\det R}}dx.$$

We continue:

$$\tilde{=} \int_{\mathbb{R}^n} \frac{R^{1/2}zz^TR^{1/2}}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}z^Tz}dz =$$

$$R^{1/2}(\int_{\mathbb{R}^n} \frac{zz^T}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}z^Tz}dz)R^{1/2} = R^{1/2}IR^{1/2} = R$$

provided

$$\int_{\mathbb{R}^n} \frac{zz^T}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}z^Tz}dz = I, \tag{3}$$

where I is the identity matrix.

To show (3), we have to verify two things:

1) all entries $(i, j), i \neq j$ in the matrix on the left side in (3) are equal to zero;

2) all entries on the diagonal in the matrix on the left side in (3) are equal to one.

For $i \neq j$, we have:

$$\int_{\mathbb{R}^n} \frac{z_i z_j}{(2\pi)^{n/2}} \exp\{-\frac{1}{2}\sum_{k=1}^{n} z_k^2\} dz =$$

$$\left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} z_i e^{-\frac{1}{2}z_i^2} dz_i\right)\left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} z_j e^{-\frac{1}{2}z_j^2} dz_j\right) \prod_{k \neq i,j} \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_k^2} dz_k\right) =$$

$$0 \times 0 \times 1 = 0$$

since the first two integrals are equal to zero and the last term is equal to 1 (Why?).

For $i = j$, we have:

$$\left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} z_i^2 e^{-\frac{1}{2}z_i^2} dz_i\right) \prod_{k \neq i} \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_k^2} dz_k\right) =$$

$$1 \times 1 = 1$$

since both terms are equal to one (again, why?).

The verification of $E(X) = \mu$ follows similar steps using the same change of variables (Exercise).

Example. (Bivariate Gaussian distribution) We consider a Gaussian vector with $n = 2$: $X = (X_1, X_2)$ so that $\mu = (\mu_1, \mu_2) = (E(X_1), E(X_2))$ and the covariance 2×2 matrix R has the entries

$$R_{11} = \sigma_1^2, R_{22} = \sigma_2^2, R_{12} = R_{21} = \rho\sigma_1\sigma_2,$$

where $\sigma_i = Var(X_i), i = 1, 2$ are the variances and ρ is the correlation coefficient between X_1 and X_2.

One easily sees then that $\det R = \sigma_1^2\sigma_2^2(1-\rho^2)$ and by direct calculation we obtain

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times$$

$$\exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}]\} \quad (4)$$

for all $x_1, x_2 \in \mathbb{R}$ and $|\rho| < 1$.

The joint density of the form (4) is *the bivariate Gaussian density.* We also note that it exists for all $|\rho| < 1$. The case of $\rho = \mp 1$ is the singular case when the vector (X_1, X_2) does not have the joint density since X_1 and X_2 are then linearly dependent.

We also note that if $\rho = 0$, then it follows from (4) that

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$$

yielding that the variables X_1 and X_2 must be independent variables. In other words, in case of Gaussian variables X_1 and X_2 "independent" is equivalent to "uncorrelated".

Proposition 2. Given (X_1, X_2) is a jointly Gaussian vector, its marginal densities are $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ are 1-dimensional Gaussian densities.

To see that the above is true, we have only to make the corresponding calculations.

To do so, it is convenient to write $f_{X_1,X_2}(x_1,x_2)$ in new notations. Define

$$\psi(x) := \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

to be the 1-dimensional density of $\mathcal{N}(0,1)$ distribution.

The function

$$\psi_\rho(u,v) := \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}[u^2 - 2\rho uv + v^2]\} \quad (5)$$

is then bivariate standard Gaussian density, that is when $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1, |\rho| < 1$ in (4).

Using $\psi_\rho(u,v)$, the general pdf in (4) can be written as

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{\sigma_1\sigma_2}\psi_\rho(\frac{x_1-\mu_1}{\sigma_1}, \frac{x_2-\mu_2}{\sigma_2}). \tag{6}$$

We observe further that for $|\rho| < 1$ we can write:

$$u^2 - 2\rho uv + v^2 = u^2(1-\rho^2) + (v-\rho u)^2$$

so that (5) becomes then

$$\psi_\rho(u,v) = \frac{e^{-u^2/2}}{\sqrt{2\pi}} \frac{e^{-\frac{1}{2(1-\rho^2)}(v-\rho u)^2}}{\sqrt{2\pi}\sqrt{1-\rho^2}}. \tag{7}$$

Combining (6) and (7), we obtain that

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{\sigma_1\sigma_2}\psi(\frac{x_1-\mu_1}{\sigma_1})\frac{1}{\sqrt{1-\rho^2}}\psi(\frac{\frac{x_2-\mu_2}{\sigma_2} - \rho\frac{x_1-\mu_1}{\sigma_1}}{\sqrt{1-\rho^2}}).$$

Now we are ready to integrate:

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_{X_1,X_2}(x_1,x_2)dx_2 =$$

$$\frac{1}{\sigma_1\sigma_2}\psi(\frac{x_1-\mu_1}{\sigma_1})\int_{\mathbb{R}} \frac{1}{\sqrt{1-\rho^2}}\psi(\frac{\frac{x_2-\mu_2}{\sigma_2} - \rho\frac{x_1-\mu_1}{\sigma_1}}{\sqrt{1-\rho^2}})dx_2 =$$

- we make the change of variables $\nu := \frac{x_2-\mu_2}{\sigma_2}, \sigma_2 d\nu = dx_2$ -

$$\frac{\sigma_2}{\sigma_1\sigma_2}\psi(\frac{x_1-\mu_1}{\sigma_1})\int_{\mathbb{R}} \frac{1}{\sqrt{1-\rho^2}}\psi(\frac{\nu - \rho\frac{x_1-\mu_1}{\sigma_1}}{\sqrt{1-\rho^2}})d\nu =$$

$$\frac{1}{\sigma_1}\psi(\frac{x_1-\mu_1}{\sigma_1}) \times 1 = \frac{1}{\sigma_1}\psi(\frac{x_1-\mu_1}{\sigma_1})$$

since

$$\int_{\mathbb{R}} \frac{1}{\sqrt{1-\rho^2}}\psi(\frac{\nu - \rho\frac{x_1-\mu_1}{\sigma_1}}{\sqrt{1-\rho^2}})d\nu = 1$$

(Why?)

Thus, we have shown that X_1 has $\mathcal{N}(\mu_1, \sigma_1^2)$ distribution.

The fact that X_2 has $\mathcal{N}(\mu_2, \sigma_2^2)$ distribution is shown in a similar way.

Remark. This remark relates to a particular problem that was given to students as part of an exam (or HW).

We recall the problem: Given X and Y are i.i.d. random variables with uniform distribution $U(0,1)$, calculate $E|X-Y|$.

• The easiest way to solve the problem is to use the general knowledge that

$$E[g(X,Y)] = \int\int g(x,y) f_{X,Y}(x,y)dxdy,$$

where $g(x,y)$ is a given real-valued function.

Given that $f_{X,Y}(x,y) = 1, 0 < x < 1, 0 < y < 1$, one has then that

$$E[|X-Y|] = \int_0^1 \int_0^1 |x-y|dxdy,$$

where the last integral is easy to calculate.

• One can try first to find the pdf of the random variable $Z := X - Y$ or $Z := |X-Y|$ and then find expectation but this would be certainly a lot of additional work.

• One can use the conditioning on another random variable. This might be also not the shortest solution but to illustrate the conditioning formula, here is the solution.

$$E|X-Y| = E[(X-Y)\mathbf{1}_{\{X>Y\}}] + E[(Y-X)\mathbf{1}_{\{Y>X\}}].$$

By symmetry, it is enough to calculate one term on the right hand side.

$$I := E[(X-Y)\mathbf{1}_{\{X>Y\}}] = \int_0^1 E[(X-t)\mathbf{1}_{\{X>t\}}|Y=t]f_Y(t)dt =$$

$$\int_0^1 E[(X-t)\mathbf{1}_{\{X>t\}}]dt,$$

where we used independence of X and Y and the fact that $f_Y(t) = 1$.

Clearly, $E[(X-t)\mathbf{1}_{\{X>t\}}] = \int_0^1 (x-t)\mathbf{1}_{\{x>t\}}(x)dx = \int_t^1 (x-t)dx = 1/2 - t + t^2/2$ so that

$$I = \int_0^1 (1/2 - t + t^2/2)dt = \frac{1}{6}.$$