ESE 520 Probability and Stochastic Processes

Lecture 6

"Random vectors and their probability distributions. Multidimensional cdf's and pdf's".

Now, we can extend what we said about a single random variable to a random vector.

∘ Formally, under a *random vector* we understand a map

$$X = (X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^n,$$

where each $X_i : \Omega \to \mathbb{R}, i = 1, 2, .., n$ is random variable.

∘ An equivalent definition would be to say that X is a *measurable map* from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, that is for every Borel set $B \in \mathcal{B}(\mathbb{R}^n)$:

$$X^{-1}(B) \in \mathcal{F}.$$

∘ The probability measure

$$P_X(B) := P(w \in \Omega : X(w) \in B)$$

is then called the *probability distribution of* X.

However, similarly as in the 1-dimensional case ($n = 1$), it is then enough to know $P_X(B)$ only for a "specific collection" of sets $B \in \mathcal{B}(\mathbb{R}^n)$ leading to the introduction of the function

$$F_X(x_1, x_2, ..., x_n) := P(w : X_1(w) \leq x_1, ..., X_n(w) \leq x_n), \tag{1}$$

where $(x_1, x_2, ..., x_n) \in \mathbb{R}^n$.

In other words, $F(x_1, x_2, ..., x_n)$ is obtained if we restrict the distribution $P_X(\bullet)$ onto the collection of sets $\mathcal{A} = \{(-\infty, x_1] \times ....(-\infty, x_n], x_i \in \mathbb{R}, i = 1, 2, ..., n\}$ which generates the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$.

The function in (1) is then called the *joint cumulative distribution function* (cdf) of the vector X and it completely describes the distribution of X. Again, that fact follows then from the Caratheodory Extension Theorem.

In summary, when working with a random vector X, we can obtain everything what we possibly can be interested in when working with the cdf $F_X(x_1, x_2, ..., x_n)$ of X.

Given the cdf $F_X(x_1, x_2, ..., x_n)$, we always can find the 1-dimensional (or, more generally, of any dimension less than n) cdfs $F_{X_i}(x_i), i = 1, 2, ..., n$ (also called *marginal cdfs*) as

$$F_{X_i}(x_i) = P(X_i \leq x_i) = P(X_1 \in \mathbb{R}, ..., X_i \leq x_i, ..., X_n \in \mathbb{R}) =$$

$$F_X(\infty, ..., x_i, ..., \infty).$$

In other words, the marginal cdfs can be always obtained from the joint cdf of the vector $X = (X_1, ..., X_n)$.

When studying a random vector $X = (X_1, ..., X_n)$, a particular important case is the case when random variables $X_1, X_2, ..., X_n$ are independent.

Definition 1. Random variables X_1 and X_2 are said to be *independent* if they assume their values independently. That is, if for all Borel sets B_1 and B_2, it holds

$$P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2). \tag{2}$$

Clearly, for $B_1 = (-\infty, x_1]$ and $B_2 = (-\infty, x_2]$, the relation becomes

$$F_X(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2). \tag{3}$$

The relations (2) and (3) can be then extended for any number $n = 2, 3, ...$.

Also, one can easily prove then the following statement:
random variables $X_1, X_2, ..., X_n$ are independent if and only if

$$F_X(x_1, x_2, .., x_n) = F_{X_1}(x_1)...F_{X_n}(x_n). \tag{4}$$

As in the case of $n = 1$, there is one particularly important class of vectors.

Definition 2. A random vector $X = (X_1, X_2, ..., X_n)$ is said to be *jointly continuous* if there is a function $f : \mathbb{R}^n \to \mathbb{R}$ such that
i) $f(x) \geq 0$ for all $x \in \mathbb{R}^n$;
ii) $\int_{\mathbb{R}^n} f(x)dx = 1$;
and

$$P(w : X(w) \in B) = \int_B f(x)dx \tag{5}$$

for all Borel sets $B \in \mathcal{B}(\mathbb{R}^n)$.

In particular, (5) yields that

$$F_X(x_1, ..., x_n) = \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} f(y_1, ...y_n)dy_n...dy_1. \tag{6}$$

The function $f(x_1, ..., x_n)$ is then called *the joint pdf* of the vector X.

It follows that if the vector $X = (X_1, ..., X_n)$ is jointly continuous, then each X_i is a continuous random variable.

Indeed: For a fixed $i = 1, 2, ..., n$, we can write that

$$F_{X_i}(x_i) = F_X(\infty, ..., x_i, ..., \infty) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{x_i} .... \int_{-\infty}^{\infty} f(y_1, ...y_n)dy_n...dy_1$$

and after differentiating the last relation

$$f_{X_i}(x_i) = F'_{X_i}(x_i) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(y_1, ..., x_i, ..., y_n)dy_n...dy_{i+1}dy_{i-1}.....dy_1. \tag{7}$$

It shows that $f_{X_i}(x_i)$ exists and can be found from the joint pdf of the vector by integrating the joint density in all other variables except x_i as (7) indicates.

In case of a jointly continuous vector the relation (4) becomes then: *random variables $X_1, X_2, ..., X_n$ are independent if and only if*

$$f_X(x_1, x_2, .., x_n) = f_{X_1}(x_1)...f_{X_n}(x_n). \tag{8}$$

Finally, we provide an important example of a jointly continuous vector.

Definition 3. A random vector $X = (X_1, ..., X_n)$ is said to have a *(non-degenerate) jointly continuous Gaussian distribution* if there exist a vector $\mu \in \mathbb{R}^n$ and a *positive definite* matrix R of size $n \times n$ such that the joint pdf of X is given by

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det R}} \exp\{-\frac{1}{2}(x-\mu)^T R^{-1}(x-\mu)\}. \tag{9}$$

where $x = (x_1, ..., x_n) \in \mathbb{R}^n$.

Some comments to notations here:

· vectors $x \in \mathbb{R}^n$ are understood as column-vectors not row-vectors;
· R^{-1} stands for the inverse matrix of R which exists because of R being positive definite;
· R being positive definite means that $x^T Rx > 0$ for all $x \in \mathbb{R}^n$ and $x \neq 0$;
· x^T stands for the transpose of x;

A simple but important example of a jointly Gaussian vector is a vector obtained as follows:

Let $X_i, i = 1, 2, ..., n$ be independent random variables each of which has $\mathcal{N}(\mu_i, \sigma_i^2)$ distribution.
Then the vector $X = (X_1, ..., X_n)$ has the joint pdf

$$f_X(x) = f_{X_1}(x_1)...f_{X_n}(x_n) =$$

$$\frac{1}{\sqrt{2\pi}\sigma_1} e^{-1/2(x_1-\mu_1)\frac{1}{\sigma_1^2}(x_1-\mu_1)} \cdots \frac{1}{\sqrt{2\pi}\sigma_n} e^{-1/2(x_n-\mu_n)\frac{1}{\sigma_n^2}(x_n-\mu_n)} =$$

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

where $\mu = (\mu_1, ..., \mu_n)$ and R is a diagonal matrix with the diagonal elements σ_i^2.

Thus, a random vector $X = (X_1, ..., X_n)$ whose components X_i are independent Gaussian random variables is a jointly continuous Gaussian vector.

It rurns out that the converse is also true: if a jointly continuous random vector $X = (X_1, ..., X_n)$ has the pdf of the form (9), where R is a diagonal matrix as described above, then the random variables $X_1, ..., X_n$ are independent Gaussian random variables.