Math 464/564, O. Knill, **April 1997**

Multidimensional distributions (Summary I)

DEFINITION. Given n random variables $X_1, \dots, X_n$. The map $X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ is called a **random vector**.

DEFINITION. The **joint distribution function** of $X = (X_1, X_2, \dots, X_n)$ is the function F on $\mathbb{R}^n$:

$$F(x_1, \dots, x_n) = \mathrm{P}[X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n] \,.$$

DEFINITION. Like in the one dimensional case, a distribution function is called **discrete** if there exists a countable set $C \subset \mathbb{R}^d$ such that $\mathrm{P}[X \in C] = 1$. It is called **absolutely continuous** if there exists a function $f : \mathbb{R}^n \to \mathbb{R}^+$ such that

$$F(t_1, \dots t_n) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_n} f(y_1, y_2, \dots, y_n) \, dy \,.$$

The function $f : \mathbb{R}^n \to \mathbb{R}$ is called the **joint density**.

LEMMA. One can recover a probability space $(\mathbb{R}^n, \mathcal{A}, \mathrm{P})$ given a joint distribution function F. One has for example in the case $n = 2$ the formula $\mathrm{P}[(a, b] \times (c, d]] = F(b, d) + F(a, c) - F(a, d) - F(a, c)$.

LEMMA. If F is a absolutely continuous distribution function. The density is

$$f(x) = F_{12,\dots,n}(x) = \frac{\partial}{\partial x_1} \dots \frac{\partial}{\partial x_n} F(x) \,.$$

For $n = 2$ one has especially $f(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} F(x, y)$.

DEFINITION. Given a joint distribution function F of the random vector, then $F_{X_i}(x) = F(\infty, \infty, \dots, x, \dots, \infty)$ is called the $i'th$ **marginal distribution** of X. It is the distribution function of the random variable X_i because $\mathrm{P}[X_i \le x, X_j \le \infty, \forall j \ne i] = \mathrm{P}[X_i \le x]$.

REMARK. Given an absolutely continuous random vector with joint density f, then the density of X_i is

$$f_{X_i}(x) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{x} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d) \, dx_1, \dots dx_n \,.$$

(Integrate over all variables except over the variable x_i). For example, for $n = 2$ one has for the random vector (X, Y) the formulas $F_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy$ and $F_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx$.

DEFINITION. Two random variables are called **independent** if

$$(*) \qquad \mathrm{P}[X \in [a, b], Y \in [c, d]] = \mathrm{P}[X \in [a, b]] \cdot \mathrm{P}[Y \in [c, d]] \,.$$

THEOREM. X and Y are independent if and only if

$$(**) \qquad F(x, y) = F_X(x) F_Y(y) \,.$$

(**) follows from (*) for $a = c = -\infty$. Using $F(a, b) = F(-\infty, b) - F(-\infty, a)$ and $F(c, d) = F(-\infty, d) - F(-\infty, c)$, one obtains the other direction $(**) \Rightarrow (*)$.

COROLLARY. Assume X and Y are absolutely continuous. By differentiation of (**), one obtains the important fact:
X and Y are independent if and only if $f(x, y) = f_X(x) f_Y(y)$.