Lectures 13-14

Definition 6.1. The σ-field generated by all open sets in $\mathbb{R}^n$ is called the Borel σ-field of subsets of $\mathbb{R}^n$ and is denoted by $\mathcal{B}_{\mathbb{R}^n}$.

Some Borel sets : Rectangles, triangles, lines, points all are Borel sets in $\mathbb{R}^2$.

Theorem 0.1 *Let*

$$\mathcal{I}_n \;=\; \big\{(-\infty,\ x_1] \times \cdots \times (-\infty,\ x_n] \,|\, (x_1, \ldots, x_n) \in \mathbb{R}^n\big\}\,.$$

Then

$$\sigma(\mathcal{I}_n) \;=\; \mathcal{B}_{\mathbb{R}^n}\,.$$

Proof. (Reading exercise) We prove for $n = 2$, for $n \geq 3$, it is similar. Note that

$$\mathcal{I}_2 \;\subseteq\; \mathcal{B}_{\mathbb{R}^2}\,.$$

Hence from the definition of $\sigma(\mathcal{I}_2)$, we have

$$\sigma(\mathcal{I}_2) \;\subseteq\; \mathcal{B}_{\mathbb{R}^2}\,.$$

Note that for $(x_1, x_2) \in \mathbb{R}^2$,

$$(-\infty,\ x_1) \times (-\infty,\ x_2) \;=\; \bigcup_{m=1}^{\infty} \Big[\Big(-\infty,\ x_1 - \frac{1}{m}\Big] \times \Big(-\infty,\ x_2 - \frac{1}{m}\Big]\Big] \in \sigma(\mathcal{I}_2)\,.$$

For each $x_1, x_2, y_1, y_2 \in \mathbb{R}$ such that $x_1 < y_1, x_2 < y_2$ we have

$$\begin{aligned}
(x_1,\ y_1) \times (x_2,\ y_2) \;=\;& (-\infty,\ y_1) \times (-\infty,\ y_2) \\
& \setminus \Big[(-\infty,\ x_1] \times (-\infty,\ x_2] \cup (-\infty,\ x_1] \times (-\infty,\ y_2] \\
& \cup (-\infty,\ y_1] \times (-\infty,\ x_2]\Big]\,.
\end{aligned}$$

Hence all bounded open rectangles are in $\sigma(\mathcal{I}_2)$. Since any open set in $\mathbb{R}^2$ can be rewritten as a countable union of bounded open rectangles, all open sets are in $\sigma(\mathcal{I}_2)$. Therefore from the definition of $\mathcal{B}_{\mathbb{R}^2}$, we get

$$\mathcal{B}_{\mathbb{R}^2} \;\subseteq\; \sigma(\mathcal{I}_2)\,.$$

This completes the proof. (It is advised that student try to write down the proof for $n = 3$)

Definition 6.2. Let Ω be a non empty set (sample space) and $\mathcal{F}$ be a σ-field of subsets of Ω. A map $X : \Omega \to \mathbb{R}^n$, is called a random vector (with respect to $\mathcal{F}$) if

$$X^{-1}(B) \in \mathcal{F} \text{ for all } B \in \mathcal{B}_{\mathbb{R}^n} .$$

Theorem 0.2 *X is a random vector with respect to $\mathcal{F}$ iff*

$$\Big\{X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n\Big\} \in \mathcal{F} \text{ for all } x_1, x_2, \cdots, x_n \in \mathbb{R}.$$

Proof: if part follows from the definition. We prove the only if part.

Suppose $X : \Omega \to \mathbb{R}^n$ satisfies

$$\Big\{X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n\Big\} \in \mathcal{F} \text{ for all } x_1, x_2, \cdots, x_n \in \mathbb{R}.$$

Let

$$\mathcal{D} \;=\; \{B \in \mathcal{B}_{\mathbb{R}^n} | X^{-1}(B) \in \mathcal{F}\}.$$

Clearly

$$\mathcal{I}_n \subseteq \mathcal{D}.$$

Now it is easy to check that $\mathcal{D}$ is a σ-field (exercise). Hence

$$\sigma(\mathcal{I}_n) \subseteq \mathcal{D} \subseteq \mathcal{B}_{\mathbb{R}^n}.$$

Therefore using Theorem 0.1, we have $\mathcal{D} = \mathcal{B}_{\mathbb{R}^n}$. i.e. for eaach $B \in \mathcal{B}_{\mathbb{R}^n}$, $X^{-1}(B) \in \mathcal{F}$. This completes the proof.

Now we introduce Borel functions in the current context.

Definition 0.1 *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be a Borel function if*

$$f^{-1}(B) \in \mathcal{B}_{\mathbb{R}^n}, \ \forall \ B \in \mathcal{B}_{\mathbb{R}^m}.$$

Lemma 0.1 *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous, then it is a Borel function.*

Proof: The main ingradient of the proof is the following fact. $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous iff $f^{-1}(O)$ is open for each open set $O \subseteq \mathbb{R}^m$.

Using this, the proof is standard, so is left as an exercise. □

Example 0.1

1. *$f_1(x, y) = xy$, $f_2(x, y) = x + y$ are Borel functions, since they are continuous functions.*

2. *Define $\pi_i : \mathbb{R}^n \to \mathbb{R}$ as follows: for each $i = 1, 2, \cdots, n$,*

$$\pi_i(x_1, x_2, \cdots, x_n) = x_i,\ (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n.$$

Then π_i is the ith coordinate projection map and is continuous and hence is a Borel function.

Exercise 6.1 Let B is a Borel set in $\mathbb{R}$, then show that $B \times \mathbb{R}$ is a Borel set in $\mathbb{R}^2$.

Method 1: Use the following argument. Collect all Borel set B satisfying "$B \times \mathbb{R}$ is Borel set in $\mathbb{R}^2$'. Then show that it is a σ-field. Also show that this collection contains all open sets in $\mathbb{R}$. Now show that (exercise) the collection is indeed all Borel sets in $\mathbb{R}^2$.

Method 2: Note that (exercise) $\pi_1^{-1}(B) = B \times \mathbb{R}$. Since π_1 is a Borel function, it follows that $B \times \mathbb{R} = \pi_1^{-1}(B) \in \mathcal{B}_{\mathbb{R}^2}$ for $B \in \mathcal{B}_{\mathbb{R}}$.

Theorem 0.3 *$X : \Omega \to \mathbb{R}^n$ is a random vector iff X_i, $i = 1, 2, \cdots, n$ are random variables where X_i denote the i^{th} component of X.*

Proof: Let X be a random vector.

For $B \in \mathcal{B}_{\mathbb{R}}$

$$X_1^{-1}(B) \;=\; X^{-1}(B \times \overbrace{\mathbb{R} \times \cdots \times \mathbb{R}}^{(n-1)\ \text{terms}}) \in \mathcal{F}\,,$$

since $B \in \mathcal{B}_{\mathbb{R}}$ we have $B \times \overbrace{\mathbb{R} \times \cdots \times \mathbb{R}}^{(n-1)\ \text{terms}} \in \mathcal{B}_{\mathbb{R}^n}$. (exercise, hint: Exercise 6.1) Therefore X_1 is a random variable. Similarly, we can show that X_i is a random variable for $i \geq 2$.

Suppose $X_1, X_2, \cdots, X_n$ are random variables.

For $x_1, x_2, \cdots, x_n \in \mathbb{R}$,

$$\big\{X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n\big\} = \big\{X_1 \leq x_1\big\} \cap \big\{X_2 \leq x_2\big\} \cap \cdots \cap \big\{X_n \leq x_n\big\} \in \mathcal{F}.$$

This completes the proof. $\square$

Theorem 0.4 *Let $X = (X_1, X_2, \cdots, X_n)$ be a random vector defined on a probability space $(\Omega, \mathcal{F}, P)$. On $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ define μ as follows*

$$\mu(B) \;=\; P\{X \in B\}\,.$$

Then μ is a probability measure on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$.

Proof. Since $\{X \in \mathbb{R}^n\} = \Omega$, we have

$$\mu(\mathbb{R}^n) \; = \; 1\,.$$

Let $B_1, B_2 \ldots$ be pair wise disjoint elements from $\mathcal{B}_{\mathbb{R}^n}$. Then $X^{-1}(B_1), X^{-1}(B_2), \ldots$ are pair wise disjoint and are in $\mathcal{F}$. Hence

$$\mu\Big(\bigcup_{m=1}^{\infty} B_m\Big) \; = \; P\Big(\bigcup_{m=1}^{\infty} X^{-1}(B_m)\Big) \; = \; \sum_{m=1}^{\infty} P\big(X \in B_m\big) \; = \; \sum_{m=1}^{\infty} \mu(B_m)\,.$$

This completes the proof. $\square$

Definition 6.3. The probability measure μ given in Theorem 0.4 is called the distribution (or Law) of the random vector X and is denoted by μ_X.

Definition 6.4. (joint distribution function)
Let $X \; = \; (X_1, \, X_2, \cdots, X_n)$ be a random vector. Then the function $F : \mathbb{R}^n \to \mathbb{R}$ given by

$$F(x_1, \, x_2, \cdots, x_n) \; = \; P\big(X_1 \le x_1, \, X_2 \le x_2, \cdots, X_n \le x_n\big)$$

is called the distribution function of X (in otherwords, joint distribution function of the random variables $X_1, X_2, \cdots, X_n$.

Now onwards **we set** $n = 2$. Let $x_1 \wedge x_2$ denote the minimum and $x_1 \vee x_2$ denote the maximum of the real numbers x_1, x_2.

Theorem 0.5 *Let F be the joint distribution function of a random vector X. Then F satisfies the following.*
(i) (a)

$$\lim_{x_1 \vee x_2 \to -\infty} F(x_1, \, x_2) \; = \; 0,$$

(b)

$$\lim_{x_1 \wedge x_2 \to \infty} F(x_1, \, x_2) \; = \; 1$$

(ii) F is right continuous in each argument, i.e.

$$\lim_{y \downarrow x} F(y, x_2) \; = \; F(x, x_2) \text{ for all } x \in \mathbb{R}, \text{ and for all } x_2 \in \mathbb{R},$$

$$\lim_{y \downarrow x} F(x_1, y) \; = \; F(x_1, x) \text{ for all } x \in \mathbb{R}, \text{ and for all } x_1 \in \mathbb{R}.$$

(iii) F is with non negative increments, i.e. given $a = (a_1, a_2), b = (b_1, b_2)$ with $a_1 < b_1, a_2 < b_2$,

$$F(\mathbf{a},\ \mathbf{b}] \geq 0,$$

where $(\mathbf{a},\ \mathbf{b}] := (a_1, b_1] \times (a_2, b_2]$ and

$$F(\mathbf{a},\ \mathbf{b}] := F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2).$$

The proof of the above theorem is an easy exercise to the student. For example to prove (iii), observe that

$$F(\mathbf{a}, \mathbf{b}] = P(X \in (\mathbf{a}, \mathbf{b}]).$$

Remark 0.1 *As in the random variable case (i.e. with $n = 1$), the above properties (i) to (iii) becomes characterizing properties for distribution function, i.e. if $F : \mathbb{R}^2 \to \mathbb{R}$ satisfies (i) to (iii), then there exists a random vector X such that F is the distribution function of X.*

Given a random vector $X = (X_1, X_2)$, the distribution function of X_1 denoted by F_{X_1} is called the marginal distribution of X_1. Similarly the marginal distribution function F_{X_2} of X_2 is defined. Given the joint distribution function F of X, one can recover the corresponding marginal distributions as follows.

$$F_{X_1}(x_1) = P\{X_1 \leq x_1\} = P\{X_1 \leq x_1,\ X_2 \in \mathbb{R}\} = \lim_{x_2 \to \infty} F(x_1,\ x_2)\,.$$

Similarly

$$F_{X_2}(x_2) = \lim_{x_1 \to \infty} F(x_1,\ x_2).$$

Given the marginal distribution functions of X_1 and X_2, in general it is impossible to construct the joint distribution function. Note that marginal distribution functions doesn't contain information about the dependence of X_1 over X_2 and vice versa. One can characterize the independence of X_1 and X_2 in terms of its joint and marginal distribution functions as in the following theorem. The proof is beyond the scope of this course.

Theorem 0.6 *Let $X = (X_1, X_2)$ be a random vector with distribution function F. Then X_1 and X_2 are independent iff*

$$F(x_1, x_2) = F_{X_1}(x_1)\, F_{X_2}(x_2),\ x_1, x_2 \in \mathbb{R}\,.$$