Lecture-05: Random Vectors

1 Random vectors

Definition 1.1 (Projection). For a vector $x \in \mathbb{R}^n$, we can define $\pi_i : \mathbb{R}^n \to \mathbb{R}$ is the **projection** of an n-length vector onto its i-th component, such that $\pi_i(x) = x_i$.

Definition 1.2. The Borel sigma algebra over space $\mathbb{R}^n$ is defined as the smallest sigma algebra generated by the family $(\pi_i^{-1}(B_x) : x \in \mathbb{R}, i \in [n])$ and is denoted by $\mathcal{B}(\mathbb{R}^n)$. The elements of the Borel sigma algebra are called Borel sets.

Remark 1. By definition of $\mathcal{B}(\mathbb{R}^n)$, projection $\pi_i : \mathbb{R}^n \to \mathbb{R}$ is a Borel measurable function for all $i \in [n]$. For a subset $A \subseteq \mathbb{R}$ and projection $\pi_i : \mathbb{R}^n \to \mathbb{R}$, we can write

$$\pi_i^{-1}(A) = \{x \in \mathbb{R}^n : x_i \in A\} = \mathbb{R} \times \cdots \times A \times \cdots \times \mathbb{R}.$$

Thus, for any $A \in \mathcal{B}(\mathbb{R})$, we have $\pi_i^{-1}(A) \in \mathcal{B}(\mathbb{R}^n)$.

Definition 1.3 (Random vectors). Consider a probability space $(\Omega, \mathcal{F}, P)$ and a finite $n \in \mathbb{N}$. A **random vector** $X : \Omega \to \mathbb{R}^n$ is an $\mathcal{F}$-measurable mapping from the sample space to an n-length real-valued vector. That is, for any $x \in \mathbb{R}^n$, we have

$$A_X(x) \triangleq \{\omega \in \Omega : X_1(\omega) \leqslant x_1, \dots, X_n(\omega) \leqslant x_n\} = \cap_{i=1}^n X_i^{-1}(-\infty, x_i] \in \mathcal{F}.$$

Example 1.4 (Tuple of indicators). Consider a probability space $(\Omega, \mathcal{F}, P)$, a finite $n \in \mathbb{N}$, and events $A_1, \dots, A_n \in \mathcal{F}$. We define a mapping $X : \Omega \to \{0,1\}^n$ by $X_i(\omega) \triangleq \mathbb{1}_{A_i}(\omega)$ for all outcomes $\omega \in \Omega$. Let $x \in \mathbb{R}^n$, then we can write $A_X(x) = \cap_{i=1}^n \mathbb{1}_{A_i}^{-1}(-\infty, x_i]$. Recall that

$$\mathbb{1}_{A_i}^{-1}(-\infty, x_i] = \begin{cases} \Omega, & x_i \geqslant 1, \\ A_i^c, & x_i \in [0,1), \\ \emptyset, & x_i < 0. \end{cases}$$

It follows that the inverse image $A_X(x)$ lies in $\mathcal{F}$, and hence X is an $\mathcal{F}$-measurable random vector.

Theorem 1.5. *Consider a probability space $(\Omega, \mathcal{F}, P)$, and a finite $n \in \mathbb{N}$. A mapping $X : \Omega \to \mathbb{R}^n$ is a random vector if and only if $X_i \triangleq \pi_i \circ X : \Omega \to \mathbb{R}$ are random variables for all $i \in [n]$.*

Proof. We will first show that $X : \Omega \to \mathbb{R}^n$ implies that $\pi_i \circ X$ is a random variable for any $i \in [n]$. For any $i \in [n]$ and $x_i \in \mathbb{R}$, we take $x = (\infty, \dots, x_i, \dots, \infty)$. This implies that $\pi_i^{-1}(-\infty, x_i] = \mathbb{R} \times \cdots \times (-\infty, x_i] \times \cdots \times \mathbb{R} \in \mathcal{B}(\mathbb{R}^n)$. Further, defining $A_{X_i}(x_i) \triangleq X_i^{-1}(-\infty, x_i]$, we observe from the definition of random vectors that

$$A_X(x) = \cap_{j=1}^n X_j^{-1}(-\infty, x_j] = X_i^{-1}(-\infty, x_i] = A_{X_i}(x_i) \in \mathcal{F}. \quad (1)$$

We will next show that if $X_i : \Omega \to \mathbb{R}$ is a random variable for all $i \in [n]$, then $X \triangleq (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ is a random vector. For any $x \in \mathbb{R}^n$, we have $A_{X_i}(x_i) = X_i^{-1}(-\infty, x_i] \in \mathcal{F}$ for all $i \in [n]$, from the definition of random variables. From the closure of event set under countable intersections, we have

$$A_X(x) = \cap_{i=1}^n A_{X_i}(x_i) \in \mathcal{F}. \quad (2)$$

□

1.1 Distribution of random vectors

Definition 1.6. Consider a probability space $(\Omega, \mathcal{F}, P)$ and a finite $n \in \mathbb{N}$. The **joint distribution function** of a random vector $X : \Omega \to \mathbb{R}^n$ is defined as the mapping $F_X : \mathbb{R}^n \to [0,1]$ such that

$$F_X(x) \triangleq P(A_X(x)) = P(\cap_{i=1}^n A_{X_i}(x_i)).$$

Example 1.7 (Tuple of indicators). Consider a probability space $(\Omega, \mathcal{F}, P)$, a finite $n \in \mathbb{N}$, and events $A_1, \dots, A_n \in \mathcal{F}$, that define the random vector $X \triangleq (\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n})$. For any $x \in \mathbb{R}^n$, we can define index sets $I_0(x) \triangleq \{i \in [n] : x_i < 0\}$ and $I_1(x) \triangleq \{i \in [n] : x_i \in [0,1)\}$, and write the joint distribution function for this random vector X as

$$F_X(x) = \begin{cases} 1, & I_0(x) \cup I_1(x) = \emptyset, \\ P(\cap_{i \in I_1(x)} A_i^c), & I_0(x) = \emptyset, I_1(x) \neq \emptyset, \\ 0, & I_0(x) \neq \emptyset. \end{cases}$$

Definition 1.8. For a random vector $X : \Omega \to \mathbb{R}^n$ defined on the probability space $(\Omega, \mathcal{F}, P)$ and $i \in [n]$, the distribution of the ith random variable $X_i \triangleq \pi_i \circ X : \Omega \to \mathbb{R}$ is called the ith **marginal distribution**, and denoted by $F_{X_i} : \Omega \to [0,1]$.

Example 1.9 (Tuple of indicators). Consider a probability space $(\Omega, \mathcal{F}, P)$, a finite $n \in \mathbb{N}$, and events $A_1, \dots, A_n \in \mathcal{F}$, that define the random vector $X \triangleq (\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n})$. The ith marginal distribution is given by

$$F_{X_i}(x) = (1 - P(A_i))\mathbb{1}_{[0,1)}(x) + \mathbb{1}_{[1,\infty)}(x).$$

Corollary 1.10 (Marginal distribution). *Consider a random vector $X : \Omega \to \mathbb{R}^n$ defined on a probability space $(\Omega, \mathcal{F}, P)$ with the joint distribution $F_X : \mathbb{R}^n \to [0,1]$. The ith **marginal distribution** and can be obtained from the joint distribution of X as*

$$F_{X_i}(x_i) = \lim_{x_j \to \infty, \text{ for all } j \neq i} F_X(x).$$

Proof. For any $i \in [n]$ and $x_i \in \mathbb{R}$, we have $X_i^{-1}(-\infty, x_i] = A_X(x)$ for $x = (\infty, \dots, x_i, \dots, \infty)$ from (1). □

Lemma 1.11 (Properties of the joint distribution function). *Consider a random vector $X : \Omega \to \mathbb{R}^n$ defined on the probability space $(\Omega, \mathcal{F}, P)$. The associated joint distribution function $F_X : \mathbb{R}^n \to [0,1]$ satisfies the following properties.*

(i) For $x, y \in \mathbb{R}^n$ such that $x_i \leqslant y_i$ for each $i \in [n]$, we have $F_X(x) \leqslant F_X(y)$.

(ii) The function $F_X(x)$ is right continuous at all points $x \in \mathbb{R}^n$.

(iii) The lower limit is $\lim_{x_i \to -\infty} F_X(x) = 0$, and the upper limit is $\lim_{x_i \to \infty, i \in [n]} F_X(x) = 1$.

Proof. Consider a random vector $X : \Omega \to \mathbb{R}^n$ defined on the probability space $(\Omega, \mathcal{F}, P)$ and any $x \in \mathbb{R}^n$.

(i) We can verify that $A_X(x) = \cap_{i=1}^n A_{X_i}(x_i) \subseteq \cap_{i=1}^n A_{X_i}(y_i) = A(y)$. The result follows from the monotonicity of probability measure.

(ii) The proof is similar to the proof for single random variable.

(iii) The event $A_X(x) = \emptyset$ when $x_i = -\infty$ for some $i \in [n]$ and $A_X(x) = \Omega$ when $x_i = \infty$ for all $i \in [n]$, hence the result follow.

□

Example 1.12 (Probability of rectangular events). Consider a probability space $(\Omega, \mathcal{F}, P)$ and a random vector $X : \Omega \to \mathbb{R}^2$. Consider the points $x \leqslant y \in \mathbb{R}^2$ and the events

$$B_1 \triangleq \{x_1 < X_1 \leqslant y_1\} = A_{X_1}(y_1) \setminus A_{X_1}(x_1) \in \mathcal{F}, \qquad B_2 \triangleq \{x_2 < X_2 \leqslant y_2\} = A_{X_2}(y_2) \setminus A_{X_2}(x_2) \in \mathcal{F}.$$

The marginal probabilities are given by

$$\begin{aligned}P(B_1) = P(A_{X_1}(y_1)) - P(A_{X_1}(x_1)) = F_{X_1}(y_1) - F_{X_1}(x_1),\\ P(B_2) = P(A_{X_2}(y_2)) - P(A_{X_2}(x_2)) = F_{X_2}(y_2) - F_{X_2}(x_2).\end{aligned}$$

Writing $x = (x_1, x_2)$ and $y = (y_1, y_2)$, we observe that the end points of the rectangular event $B_1 \cap B_2$ are points $x, (y_1, x_2), y, (x_1, y_2)$. Therefore, we can write this event as

$$B_1 \cap B_2 = (A_X(y) \setminus A_X(x_1, y_2)) \setminus (A_X(y_1, x_2) \setminus A_X(x)) \in \mathcal{F}.$$

Hence, we can write the probability of this rectangular event as

$$P(B_1 \cap B_2) = (F_X(y) - F_X(x_1, y_2)) - (F_X(y_1, x_2) - F_X(x)).$$

1.2 Event space generated by random vectors

Definition 1.13. Consider a probability space $(\Omega, \mathcal{F}, P)$ and a finite $n \in \mathbb{N}$. The **event space generated by a random vector** $X : \Omega \to \mathbb{R}^n$ is the smallest σ-algebra generated by the collection of events $(A_X(x) : x \in \mathbb{R}^n)$ and denoted by $\sigma(X) \triangleq \sigma(A_X(x) : x \in \mathbb{R}^n)$.

Theorem 1.14. *Consider a probability space $(\Omega, \mathcal{F}, P)$, a finite $n \in \mathbb{N}$, a random vector $X : \Omega \to \mathbb{R}^n$, and its projections $X_i \triangleq \pi_i \circ X$ for all $i \in [n]$. Then, $\sigma(X) = \sigma(X_1, \dots, X_n)$.*

Proof. Recall that $\sigma(X)$ is generated by the family $(A_X(x) : x \in \mathbb{R}^n)$ and $\sigma(X_1, \dots, X_n)$ is generated by the family $(A_{X_i}(x_i) : x_i \in \mathbb{R}, i \in [n])$. We first show that $A_{X_i}(x_i) = A_X(x)$ for $x = (\infty, \dots, x_i, \dots, \infty)$, and hence $\sigma(X_1, \dots, X_n) \subseteq \sigma(X)$. We then show that $A_X(x) = \cap_{i=1}^n A_{X_i}(x_i)$, and hence $\sigma(X) \subseteq \sigma(X_1, \dots, X_n)$. $\square$

Example 1.15 (Tuple of indicators). Consider a probability space $(\Omega, \mathcal{F}, P)$, a finite $n \in \mathbb{N}$, and events $A_1, \dots, A_n \in \mathcal{F}$, that define the random vector $X \triangleq (\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n})$. The $\sigma(X) = \sigma(A_i : i \in [n])$.

1.3 Independence of random variables

Definition 1.16. A family of collections of events $(\mathcal{A}_i \subseteq \mathcal{F} : i \in I)$ is called independent, if for any finite set $F \subseteq I$ and $A_i \in \mathcal{A}_i$ for all $i \in F$, we have

$$P(\cap_{i \in F} A_i) = \prod_{i \in F} P(A_i).$$

Definition 1.17 (Independent and identically distributed). A random vector $X : \Omega \to \mathbb{R}^n$ defined on the probability space $(\Omega, \mathcal{F}, P)$ is called **independent** if

$$F_X(x) = \prod_{i=1}^n F_{X_i}(x_i), \text{ for all } x \in \mathbb{R}^n.$$

The random vector X is called **identically distributed** if each of its components have the identical marginal distribution, i.e.

$$F_{X_i} = F_{X_1}, \text{ for all } i \in [n].$$

Remark 2. Independence of a random vector implies that events $(A_{X_i}(x_i) : i \in [n])$ are independent for any $x \in \mathbb{R}^n$. Defining families $\mathcal{A}_i \triangleq (A_{X_i}(x) : x \in \mathbb{R})$ for all $i \in [n]$, we observe that the families $(\mathcal{A}_1, \dots, \mathcal{A}_n)$ are mutually independent.

Remark 3. In general, if two collection of events are mutually independent, then the event space generated by them are independent. This can be proved using Dynkin's π-λ Theorem.

Theorem 1.18. *For an independent random vector $X : \Omega \to \mathbb{R}^n$ defined on a probability space $(\Omega, \mathcal{F}, P)$, the event spaces generated by its components $(\sigma(X_i) : i \in [n])$ are independent.*

Proof. For an we define a family of events $\mathcal{A}_i \triangleq (X_i^{-1}(-\infty, x] : x \in \mathbb{R})$ for each $i \in [n]$. From the definition of independence of random vectors, the families $(\mathcal{A}_i \subseteq \mathcal{F} : i \in [n])$ are mutually independent. Since $\sigma(\mathcal{A}_i) = \sigma(X_i)$, the result follows from the previous remark. □

Definition 1.19 (Independent random vectors). To random vectors $X : \Omega \to \mathbb{R}^n$ and $Y : \Omega \to \mathbb{R}^m$ defined on the same probability space $(\Omega, \mathcal{F}, P)$ are independent, if the collection of events $(A_X(x) : x \in \mathbb{R}^n)$ and $(A_Y(y) : y \in \mathbb{R}^m)$ are independent, where $A_X(x) \triangleq \cap_{i=1}^n X_i^{-1}(-\infty, x_i]$ and $A_Y(y) \triangleq \cap_{i=1}^m Y_i^{-1}(-\infty, y_i]$.

Example 1.20 (Independent random vectors). Consider a set of vectors $\mathcal{X} = \{(0,0,1), (1,0,0)\} \subseteq \mathbb{R}^3$. Consider two independent coin tosses, such that $\Omega = \{H, T\}^2, \mathcal{F} = \mathcal{P}(\Omega)$ and $P(\omega) = p^{k_2(\omega)}(1-p)^{2-k_2(\omega)}$, where $k_2(\omega) = \sum_{i=1}^2 \mathbb{1}_{\{\omega_i = H\}}$. We define random vectors

$$X = (0,0,1)\mathbb{1}_{\{\omega_1 = H\}} + (1,0,0)\mathbb{1}_{\{\omega_1 = T\}}, \qquad Y = (0,0,1)\mathbb{1}_{\{\omega_2 = H\}} + (1,0,0)\mathbb{1}_{\{\omega_2 = T\}}.$$

We can verify that $X, Y : \Omega \to \mathbb{R}^3$ are mutually independent random vectors, though we can also check that X_1, X_3 are dependent random variables and so are Y_1, Y_3.

1.4 Discrete random vectors

Definition 1.21 (Discrete random vectors). If a random vector $X : \Omega \to \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \subseteq \mathbb{R}^n$ takes countable values in $\mathbb{R}^n$, then it is called a **discrete random vector**. That is, the range of random vector X is countable, and the random vector is completely specified by the **probability mass function**

$$P_X(x) = P(\cap_{i=1}^n \{X_i = x_i\}) \text{ for all } x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n.$$

Remark 4. For an independent discrete random vector $X : \Omega \to \mathbb{R}^n$, we have $P_X(x) = \prod_{i=1}^n P_{X_i}(x_i)$ for each $x \in \mathbb{R}^n$.

Example 1.22 (Multiple coin tosses). For a probability space $(\Omega, \mathcal{F}, P)$, such that $\Omega = \{H, T\}^n, \mathcal{F} = 2^\Omega, P(\omega) = \frac{1}{2^n}$ for all $\omega \in \Omega$.

Consider the random vector $X : \Omega \to \mathbb{R}$ such that $X_i(\omega) = \mathbb{1}_{\{\omega_i = H\}}$ for each $i \in [n]$. Observe that X is a bijection from the sample space to the set $\{0,1\}^n$. In particular, X is a discrete random variable.

For any $x \in [0,1]^n$, we can write $N(x) = \sum_{i=1}^n \mathbb{1}_{[0,1)}(x_i)$. Further, we can write the joint distribution as

$$F_X(x) = \begin{cases} 1, & x_i \geqslant 1 \text{ for all } i \in [n], \\ \frac{1}{2^{N(x)}}, & x_i \in [0,1] \text{ for all } i \in [n], \\ 0, & x_i < 0 \text{ for some } i \in [n]. \end{cases}$$

We can derive the marginal distribution for i-th component as

$$F_{X_i}(x_i) = \begin{cases} 1, & x_i \geqslant 1, \\ \frac{1}{2}, & x_i \in [0,1), \\ 0, & x_i < 0. \end{cases}$$

Therefore, it follows that X is an *i.i.d.* random vector.

1.5 Continuous random vectors

Definition 1.23 (Joint density function). For jointly continuous random vector $X : \Omega \to \mathbb{R}^n$ with joint distribution function $F_X : \mathbb{R}^n \to [0,1]$, there exists a **joint density function** $f_X : \mathbb{R}^n \to [0,\infty)$ such that $f_X(x) = \frac{d^n}{dx_1 \dots dx_n} F_X(x)$, and

$$F_X(x) = \int_{u_1 \leqslant x_1} du_1 \cdots \int_{u_n \leqslant x_n} du_n f_X(u_1, \dots, u_n).$$

Remark 5. For an independent continuous random vector $X : \Omega \to \mathbb{R}^n$, we have $f_X(x) = \prod_{i=1}^n f_{X_i}(x_i)$ for all $x \in \mathbb{R}^n$.

Example 1.24 (Gaussian random vectors). For a probability space $(\Omega, \mathcal{F}, P)$, Gaussian random vector is a continuous random vector $X : \Omega \to \mathbb{R}^n$ defined by its density function

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right) \text{ for all } x \in \mathbb{R}^n,$$

where the mean vector $\mu \in \mathbb{R}^n$ and the positive definite covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$. The components of the Gaussian random vector are Gaussian random variables, which are independent when Σ is diagonal matrix and are identically distributed when $\Sigma = \sigma^2 I$.