# Joint Probability Distributions, Independence

Sections 6.1, 6.2, 6.3 of ASV

*Instructor: Vincent Roulet* *Teaching Assistant: Zhenman Yuen*

## 1 Multivariate random variables

**Definition 1** (Multivariate random variable/Random vector)**.** *Given a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, a* ***multivariate random variable*** *or* ***random vector*** *is a vector* $X = (X_1, ..., X_n)$*, whose components are real-valued random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$*.*

**Example 2** (Classic examples)**.**

1. *(Discrete case) Roll a die 100 times, denote* $X_1, \ldots, X_6$ *the number of 1, ..., 6 you got respectively, then* $X = (X_1, \ldots, X_6)$ *is a random vector*
2. *(Continuous case) Throw a dart uniformly at random on a disc, the coordinates* $(X, Y)$ *of that throw form a random vector*

### 1.1 Discrete case

#### 1.1.1 Joint probability mass function

**Definition 3** (Joint probability mass function)**.** *Let* $X_1, \ldots, X_n$ *be discrete r.v. on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, their* ***joint probability mass function*** *is defined as*

$$\begin{aligned} p(k_1, \ldots, k_n) &= \mathbb{P}(\{X_1 = k_1\} \cap \ldots \cap \{X_n = k_n\}) \\ &\triangleq \mathbb{P}(X_1 = k_1, \ldots, X_k = k_n) \end{aligned}$$

*for any* $k_1, \ldots, k_n \in X_1(\Omega) \times \ldots \times X_n(\Omega)$ *(any values taken by the random vector)*

**Example 4.**

1. *Roll two dice with* ***4 faces****, denote*
   (i) *$S$ the sum of the two dice*
   (ii) *$Y$ the indicator variable that you get a pair*
2. *Record which outcomes lead to different values of $S, Y$*
3. *Compute the corresponding joint probability mass function of $S, Y$*
4. *Read e.g.* $\mathbb{P}(S = 4, Y = 1) = 1/16$

| | | Y | |
|---|---|---|---|
| | | 0 | 1 |
| | 2 | | (1,1) |
| | 3 | (1, 2) (2, 1) | |
| | 4 | (1, 3) (3, 1) | (2, 2) |
| S | 5 | (1, 4) (2, 3) (3, 2) (4, 1) | |
| | 6 | (2, 4) (4, 2) | (3, 3) |
| | 7 | (3, 4) (4, 3) | |
| | 8 | | (4, 4) |

| | | Y | |
|---|---|---|---|
| | | 0 | 1 |
| | 2 | 0 | 1/16 |
| | 3 | 1/8 | 0 |
| | 4 | 1/8 | 1/16 |
| S | 5 | 1/4 | 0 |
| | 6 | 1/8 | 1/16 |
| | 7 | 1/8 | 0 |
| | 8 | 0 | 1/16 |

**Lemma 5.** *Let $g : \mathbb{R}^n \to \mathbb{R}$ and let $X_1, \dots, X_n$ be discrete r.v. on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with joint probability mass function $p$, then*

$$\mathbb{E}[g(X_1, \dots, X_n)] = \sum_{k_1, \dots k_n \in X_1(\Omega) \times \dots \times X_n(\Omega)} g(k_1, \dots, k_n) p(k_1, \dots k_n)$$

**Example 6.** *1. Roll two dices with **4 faces**, denote*

*(i) $S$ the sum of the two dices*

*(ii) $Y$ the indicator variable that you get a pair*

*2. Score is the sum of the dice, doubled if it is a pair. What is the average score?*

*Solution.* The average score reads

$$\begin{aligned}
\mathbb{E}[g(S, Y)] &= \sum_{s=2}^{8} \sum_{y=0}^{1} s(y+1) p(s, y) \\
&= \sum_{s=2}^{8} sp(s, 0) + 2 \sum_{s=2}^{8} sp(s, 1) \\
&= \frac{3 + 4 + 2 \times 5 + 6 + 7}{8} + 2 \times \frac{2 + 4 + 6 + 8}{16} = 25/4 = 6.25
\end{aligned}$$

□

### 1.1.2 Marginal probability mass function

**Definition 7.** *Let $p_{X,Y}$ be the joint probability mass function of two r.v. $(X, Y)$. The probability mass function of $X$ is given by,*

$$p_X(k) \triangleq \mathbb{P}(X = k) = \sum_{\ell \in Y(\Omega)} p_{X,Y}(k, \ell)$$

*The function $p_X$ is called the **marginal probability distribution** of $X$.*

*Proof.* The events $\{B_\ell = \{Y = \ell\}\}_{\ell \in Y(\Omega)}$ form a partition of $\Omega$ y definition of a discrete random variable such that

$$\mathbb{P}(X = k) = \mathbb{P}\left(\{X = k\} \cap \bigcup_{\ell=-\infty}^{+\infty} B_\ell\right) = \sum_{\ell=-\infty}^{+\infty} \mathbb{P}(X = k, Y = \ell) = \sum_{\ell \in Y(\Omega)} p_{X,Y}(k, \ell)$$

□

**Definition 8.** *Let $p$ be the joint probability mass function of $n$ discrete r.v. $X_1, \dots X_n$. The probability mass function of $X_j$ for $j \in \{1, \dots, n\}$ is given by for any $k \in X_j(\Omega)$,*

$$p_{X_j}(k) = \sum_{\substack{\ell_1, \dots, \ell_{j-1}, \ell_{j+1}, \dots, \ell_n \\ \in X_1(\Omega) \times \dots \times X_{j-1}(\Omega) \times X_{j+1}(\Omega) \times \dots \times X_n(\Omega)}} p(\ell_1, \dots, \ell_{j-1}, k, \ell_{j+1}, \dots, \ell_n)$$

*The function $p_{X_j}$ is called the **marginal probability distribution** of $X_j$.*

Previous result generalizes to the joint probability distribution of any subset.
For example the joint probability of $X_1, \dots X_m$ given $m < n$ is

$$p_{X_1, \dots X_m}(k_1, \dots, k_m) = \sum_{\ell_{m+1}, \dots, \ell_n \in X_{m+1}(\Omega) \times \dots \times X_n(\Omega)} p(k_1, \dots, k_m, \ell_{m+1}, \dots \ell_n)$$

**Example 9.** *1. Roll two dices with 4 faces, denote* *(i) $S$ the sum of the two dices*

*(ii) $Y$ the indicator variable that you get a pair*

2. *Compute marginal distribution of $Y$ from the joint p.m.f.*

*Solution.* Sum the columns of $p(s, y)$. So you get $\mathbb{P}(Y = 1) = 4/16$ and $\mathbb{P}(Y = 0) = 12/16$ □

### 1.1.3 Multinomial distribution

**Motivation** Consider a trial with $r$ possible outcomes, labeled $1, \ldots, r$. Denote $p_j$ the probability of the out come $j$ such that $p_1 + \ldots + p_r = 1$. Perform $n$ independent repetitions of this trial. Denote $X_j$ the number of times the outcome $j$ appeared among the $n$ trials.

What is the joint probability mass function of $(X_1, \ldots, X_r)$?

**Derivation**

1. Let $k_1, \ldots, k_r \in \mathbb{N}$ such that $k_1 + \ldots + k_r = n$.
2. Any outcome that leads to $X_j = k_j$ for all $j \in \{1, ..., r\}$ has proba $p_1^{k_1} ... p_r^{k_r}$.
3. The number of such outcomes is given by (in book page 392)

$$\binom{n}{k_1, \ldots, k_r} = \frac{n!}{k_1! \ldots k_r!}$$

4. Therefore we get $\mathbb{P}(X_1 = k_1, \ldots, X_r = k_r) = \binom{n}{k_1, \ldots, k_r} p_1^{k_1} ... p_r^{k_r}$

**Definition 10** (Multinomial distribution)**.** *Let $n, r \in \mathbb{N}_*$, let $p_1, \ldots, p_r \in (0, 1)$ s.t. $p_1 + \ldots + p_r = 1$, then a r.v. $X$ has a* ***multinomial distribution*** *with parameters $n, r, p_1, \ldots, p_r$ if it is defined for any $k_1, \ldots k_r \in \mathbb{N}$ s.t. $k_1 + \ldots + k_r = n$ with probability*

$$\mathbb{P}(X_1 = k_1, \ldots, X_r = k_r) = \binom{n}{k_1, \ldots, k_r} p_1^{k_1} ... p_r^{k_r}$$

*We denote it $(X_1, \ldots, X_r) \sim \mathrm{Multinom}(n, r, p_1, \ldots, p_r)$.*

## 1.2 Continuous Random Variables

### 1.2.1 Joint probability density function

**Definition 11** (Joint probability density function)**.** *Random variables $X_1, \ldots, X_n$ are* ***jointly continuous*** *if there exists a* ***joint probability density function*** *$f : \mathbb{R}^n \to \mathbb{R}$ such that for any[1] $B \subset \mathbb{R}^n$,*

$$\mathbb{P}(X_1, \ldots, X_n \in B) = \int \ldots \int_B f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

$X$ and $Y$ have a p.d.f. does not imply that $(X, Y)$ is jointly continuous!

*Example:* Take $X$ any continuous r.v., define $Y = X$, s.t. $\mathbb{P}(X = Y) = 1$. If $(X, Y)$ had a joint p.d.f. $f$, denoting $D = \{(x, y) : x = y\}$, we would have

$$\mathbb{P}(X = Y) = \iint_D f(x, y) dx dy = \int_{-\infty}^{+\infty} \left( \int_x^x f(x, y) dy \right) dx = 0$$

**Lemma 12.** *Let $X_1, \ldots, X_n$ be $n$ jointly continuous r.v.. Then for any subset $A \subset \mathbb{R}^n$ included in a linear subspace $E \subset \mathbb{R}^n$ of dimension $\dim(E) = m < n$,*

$$\mathbb{P}((X_1, \ldots, X_n) \in A) = 0$$

[1] Think of $B$ as for example $[a, b]^n$. Again a rigorous definition requires $B$ to belong to the Borel algebra of $\mathbb{R}^n$

**Example 13** (Synthetic)**.** *Assume $X, Y$ have a joint p.d.f.*

$$f(x,y) = \begin{cases} \frac{3}{2}(xy^2+y) & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

1. *Check that it is a valid joint p.d.f*
2. *Compute $\mathbb{P}(X < Y)$*

*Solution.* 1. We have $f(x,y) \ge 0$ and

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = \frac{3}{2}\int_0^1 \left(\int_0^1 xy^2 + ydx\right) dy$$
$$= \frac{3}{2}\int_0^1 \left(\frac{1}{2}y^2 + y\right) dy = \frac{3}{2}\left(\frac{1}{6}+\frac{1}{2}\right) = 1$$

2.

$$\mathbb{P}(X < Y) = \frac{3}{2}\int_0^1 \left(\int_0^y (xy^2+y)dx\right) dy$$
$$= \frac{3}{2}\int_0^1 \left(\frac{1}{2}y^4 + y^2\right) dy$$
$$= \frac{3}{2}\left(\frac{1}{10}+\frac{1}{3}\right) = 0.65$$

□

### 1.2.2 Uniform continuous random variables

**Definition 14** (Uniform continuous random variable in dimension 2 or 3)**.** *Let $D$ be a bounded subset of $\mathbb{R}^2$ s.t.* $\text{Area}(D) < +\infty$. *The random point $(X,Y)$ is* ***uniformly distributed on*** *$D$ if its joint p.d.f. reads*

$$f(x,y) = \frac{1}{\text{Area}(D)}\mathbf{1}_D(x,y) = \begin{cases} \frac{1}{\text{Area}(D)} & \text{if}(x,y) \in D \\ 0 & \text{otherwise} \end{cases}$$

*Let $D$ be a bounded subset of $\mathbb{R}^3$ s.t.* $\text{Vol}(D) < +\infty$. *The random point $(X,Y,Z)$ is* ***uniformly distributed on*** *$D$ if its joint p.d.f. reads*

$$f(x,y,z) = \frac{1}{\text{Vol}(D)}\mathbf{1}_D(x,y) \begin{cases} \frac{1}{\text{Vol}(D)} & \text{if}(x,y.z) \in D \\ 0 & \text{otherwise} \end{cases}$$

*We denote $(X,Y) \sim \text{Unif}(D)$ or $(X,Y,Z) \sim \text{Unif}(D)$.*

**Lemma 15.** *Let $(X,Y) \sim \text{Unif}(D)$ for $D \subset \mathbb{R}^2$, then for any $G \subset D$, (similar for $\mathbb{R}^3$)*

$$\mathbb{P}((X,Y) \in G) = \frac{\text{Area}(G)}{\text{Area}(D)}$$

*Proof.*

$$\text{Pr}((X,Y) \in G) = \frac{1}{\text{Area}(D)}\int\int \mathbf{1}_G(x,y)\,\mathbf{1}_D(x,y)dxdy = \int\int \mathbf{1}_G(x,y)dxdy = \frac{\text{Area}(G)}{\text{Area}(D)}$$

□

### 1.2.3 Marginal Probability Density Function

**Definition 16** (Marginal probability density function)**.** *Let $X, Y$ be jointly continuous r.v. and denote $f_{X,Y}$ their joint p.d.f. then the p.d.f. of $X$ exists and is given by*

$$f_X(X) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$

*Proof.* We have by definition of the joint p.d.f. an expression of the c.d.f. of $X$ as

$$F_X(t) = \mathbb{P}(X \le t) = \mathbb{P}(X \le t, -\infty \le Y \le +\infty) = \int_{-\infty}^{t}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dydx$$

Therefore $f_X(x) = F'_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)$ $\square$

**Definition 17** (Marginal probability density function)**.** *Let $X_1, \dots X_n$ be jointly continuous and denote $f$ their joint p.d.f..*

*Then for any $j \in \{1, \dots n\}$, $X_j$ is a continuous random variable with p.d.f.*

$$f_{X_j}(x) = \underbrace{\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty}}_{(n-1\ integrals)} f(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n)dx_1 \dots dx_{j-1}dx_{j+1} \dots dx_n$$

**Example 18.** *Consider a disk of radius $r$, $D_r = \{(x,y) : x^2 + y^2 \le r\}$ and $(X,Y) \sim \mathrm{Unif}(D_r)$.*

*What is the marginal p.d.f. of $X$?*

*Solution.* Joint p.d.f. is $f_{X,Y}(x,y) = \frac{1}{\pi r^2}\mathbf{1}_{D_r}(x,y)$ where $D_r = \{(x,y) : x^2 + y^2 \le r^2\}$

Marginal density is then $f_X(x) = 0$ for $|x| > r$, and for $|x| \le r$,

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \frac{1}{\pi r^2}\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} dy = \frac{2}{\pi r^2}\sqrt{r^2 - x^2}$$

$\square$

## 1.3 Joint cumulative distribution

**Definition 19** (Joint cumulative distribution)**.** *The **joint cumulative distribution** of r.v. $X_1, \dots, X_n$ is defined as*

$$\begin{aligned} F(t_1, \dots, t_n) &= \mathbb{P}(\{X_1 \le t_1\} \cap \dots \cap \{X_n \le t_n\}) \\ &\triangleq \mathbb{P}(X_1 \le t_1, \dots, X_n \le t_n) \end{aligned}$$

**Lemma 20.**

1. *If $(X, Y)$ are jointly continuous with joint p.d.f. $f$,*

$$F(t,s) = \int_{-\infty}^{t}\int_{-\infty}^{s} f(x,y)dxdy$$

2. *If $(X, Y)$ are jointly continuous (i.e. there exists a joint p.d.f.) with joint c.d.f. $F$*

$$\frac{\partial^2}{\partial t \partial s}F(t,s)\Big|_{s=x,t=y} = f(x,y)$$

# 2 Joint Probability Distributions and Independence

## 2.1 Independence of random variables

**Definition 21** (Independent random variables)**.** *Random variables $X_1, \ldots, X_n$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are **independent** if for any*[2] *subsets $B_1, \ldots, B_n \subset \mathbb{R}$,*

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \ldots \mathbb{P}(X_n \in B_n)$$

*or equivalently if their joint c.d.f. $F$ factorizes into the marginal c.d.f. as*

$$F(t_1, \ldots, t_n) = F_{X_1}(t_1) \ldots F_{X_n}(t_n)$$

## 2.2 Discrete case

**Lemma 22.** *Let $X_1, \ldots X_n$ be $n$ discrete random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then $X_1, \ldots X_n$ are independent if and only if their joint p.m.f. $p$ factorizes into the marginals $p_{X_i}$,*

$$p(k_1, \ldots, k_n) = p_{X_1}(k_1) \ldots p_{X_n}(k_n)$$

*Proof.* If $X_1, \ldots, X_n$ are independent the result comes from the definition.
If the joint p.m.f. factorizes into the marginal distributions, then

$$\begin{aligned}
\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) &= \sum_{k_1 \in B_1, \ldots, k_n \in B_n} p(k_1, \ldots, k_n) \\
&= \sum_{k_1 \in B_1, \ldots, k_n \in B_n} p_{X_1}(k_1) \ldots p_{X_n}(k_n) \\
&= \left( \sum_{k_1 \in B_1} p_{X_1}(k_1) \right) \ldots \left( \sum_{k_n \in B_n} p_{X_n}(k_n) \right) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i)
\end{aligned}$$

□

**Example 23.**

1. *Roll two dices with **4 faces**, denote*
   - *(i) $S$ the sum of the two dices*
   - *(ii) $Y$ the indicator variable that you get a pair*
2. *Are $S, Y$ independent?*

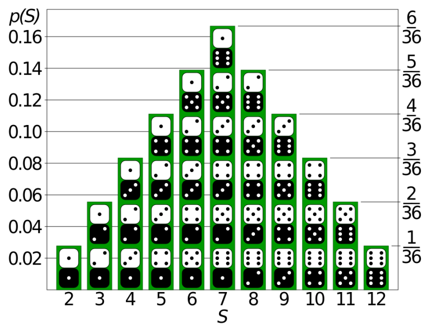
| | | *Y* | |
|---|---|---|---|
| | | *0* | *1* |
| | *2* | *0* | *1/16* |
| | *3* | *1/8* | *0* |
| | *4* | *1/8* | *1/16* |
| *S* | *5* | *1/4* | *0* |
| | *6* | *1/8* | *1/16* |
| | *7* | *1/8* | *0* |
| | *8* | *0* | *1/16* |

*Solution.* Check for example $\mathbb{P}(S = 2, Y = 0) = 0 \neq \mathbb{P}(S = 2)\,\mathbb{P}(Y = 0) > 0$
Note: one counterexample suffices to show that $S, Y$ are dependent,
but to prove independence one would need to show the equality for all values of $S, Y$ □

**Example 24.**

*Roll repeatedly a pair of dice. Denote $N$ the number of rolls until the sum of the dice is 2 or a 6*

1. *What is the distribution of $N$?*
2. *Denote $X$ the sum you finally get (2 or 6), are $X$ and $N$ independent?*




[2] Again a formal definition requires these subsets to be Borel subsets of $\mathbb{R}^n$

*Solution.* 1. Let $Y_i$ be the sum of the two dice at the $\mathrm{i}^{\mathrm{th}}$ roll.

We have $\mathbb{P}(Y_i \in \{2,6\}) = 1/36 + 5/36 = 1/6$ and so $N \sim \mathrm{Geom}(1/6)$

2. $\mathbb{P}(N = n, X = 6) = \mathbb{P}(Y_1 \notin \{2,6\}, \ldots, Y_{n-1} \notin \{2,6\}, Y_n = 6) = \left(\frac{5}{6}\right)^{n-1} \frac{5}{36}$ Therefore $\mathbb{P}(X = 6) = \sum_{n=1}^{+\infty} \left(\frac{5}{6}\right)^{n-1} \frac{5}{36} = \frac{5/36}{1-5/6} = 5/6$

So $\mathbb{P}(N = n, X = 6) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6}\frac{5}{6} = \mathbb{P}(N = n)\,\mathbb{P}(X = 6)$

Same argument shows $\mathbb{P}(N = n, X = 2) = \mathbb{P}(N = n)\,\mathbb{P}(X = 2)$

$\rightarrow N$ and $X$ are independent

□

## 2.3 Continuous case

**Lemma 25.** *Let $X_1, \ldots, X_n$ be $n$ r.v. on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that for $j \in \{1, \ldots, n\}$, the rv. $X_j$ has p.d.f. $f_{X_j}$.*

1. *If $X_1, \ldots, X_n$ have a joint p.d.f. that factorizes in the marginal p.d.f. as*

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \ldots f_{X_n}(x_n)$$

*then $X_1, \ldots, X_n$ are independent.*

2. *Conversely if $X_1, \ldots, X_n$ are independent then they are jointly continuous with joint p.d.f.*

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \ldots f_{X_n}(x_n)$$

*Proof.* For $n = 2$ with two r.v. $(X, Y)$, denote $A, B \subset \mathbb{R}$,

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) = \int_A \int_B f_{X,Y}(x,y)dxdy &= \int_A \int_B f_X(x) f_Y(y) dxdy \\ &= \int_A f_X(x)dx \int_B f_Y(y)dy = \mathbb{P}(X \in A)\,\mathbb{P}(X \in B)\end{aligned}$$

Conversely, if $X, Y$ are independent

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(X \in B) = \int_A \int_B f_X(x) f_Y(y) dxdy$$

□

**Example 26.** *Consider $X, Y$ with p.d.f. $f(x,y) = \frac{1}{\lambda} \frac{e^x}{\sqrt{y+1}} \mathbf{1}_W(x,y)$ for $\lambda{=}2(\sqrt{2}-1)(e - e^{-1})$ where $W{=}\{(x,y) : -1{\le}x{\le}1, 0{\le}y{\le}1\}$.*

1. *Are $X, Y$ independent?*

2. *What consequences it had when computing the probability to get the target $T{=}\{(x,y){:}{-0.1}{\le}x{\le}0.1, 0.4{\le}y{\le}0.6\}$?*

*Solution.* 1. Note that $\mathbf{1}_W(x,y) = \mathbf{1}_{[-1,1]}(x)\,\mathbf{1}_{[0,1]}(y)$,

then one has $f_X(x) = \frac{1}{e-e^{-1}} e^x\, \mathbf{1}_{[-1,1]}(x)$, $f_Y(y) = \frac{1}{2(\sqrt{2}-1)\sqrt{y+1}} \mathbf{1}_{[0,1]}(y)$

So $X, Y$ are independent

2. $\mathbb{P}((X,Y) \in T) = \mathbb{P}(X \in [-0.1, 0.1])\,\mathbb{P}(Y \in [0.4, 0.6])$ where $\mathbb{P}(X \in [-0.1, 0.1])$ , $\mathbb{P}(Y \in [0.4, 0.6])$ can be computed from $f_X$, $f_Y$ respectively.

□

# 3 Borel Algebra*

Until now, we defined proba. distributions on any $B \subset \mathbb{R}^n$ for $n$=1 or $n$>1. Formal definitions require to restrict our focus to subsets $B \subset \mathbb{R}^n$ that form a $\sigma$-algebra $\mathcal{B}$

**Definition 27** ($\sigma$-algebra)**.** *Let $\Omega$ be a set, a $\sigma$-algebra $\mathcal{F}$ on $\Omega$ is a subset of $2^\Omega = \{B \subset \Omega\}$ such that*

1. $\Omega \in \mathcal{F}$
2. *(Stable by complementarity) For any $A \in \mathcal{F}$, $A^c \triangleq \Omega \setminus A \in \mathcal{F}$*
3. *(Stable by countable union) For any $A_1, A_2, \ldots \in \mathcal{F}$, $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$*

**Why introducing $\sigma$-algebra?**
You want the probability measure to satisfy that

- the measure is non-negative
- the measure of the union of disjoint sets is the sum of the measure of union sets

Then you can build a union of sets $V_k$ (see e.g. Vitali set on Wikipedia) s.t.

$$[0,1] \subset \bigcup_{k=1}^{+\infty} V_k \subset [-1,2] \qquad \mathbb{P}(V_k) = \lambda \geq 0 \quad \text{for all} \quad k$$

which leads to $1 \leq \sum_{k=1}^{+\infty} \mathbb{P}(V_k) \leq 3$ which is impossible
Formally, we restrict our focus on the Borel algebra of $\mathbb{R}^n$

**Definition 28** (Borel algebra in $\mathbb{R}^n$)**.** *The Borel algebra in $\mathbb{R}^n$, denoted $\mathcal{B}_n$, is the smallest $\sigma$-algebra (in terms of inclusion) that contains*

- *all product of intervals $[a_1, b_1] \times \ldots \times [a_n, b_n]$ for $a_i \leq b_i \in \mathbb{R}$*

*or equivalently defined as the smallest $\sigma$-algebra that contains*

- *all product of intervals of the form $(-\infty, a_1] \times \ldots \times (-\infty, a_n]$ for $a_i \in \mathbb{R}$.*

**Consequence**

1. If we can measure all intervals of the form $(-\infty, a_1] \times \ldots \times (-\infty, a_n]$ for $a_i \in \mathbb{R}$, then we can measure all subsets of interests, i.e. all $B \in \mathcal{B}_n$,

   $\rightarrow$ we know all the information necessary to describe the proba distribution

2. All the information necessary to describe any r.v. is contained in its c.d.f.