# Probability and Statistics 1 Notes

Francesco Chotuck

## Contents

# 1 Elementary probability theory

## 1.1 Axiomatic approach to probability

**Definition 1.1.** An **experiment** is the process by which observations are made.

**Definition 1.2.** The outcome of an individual observation is called an **elementary event**, we associate points called **sample points** to each elementary event.

**Definition 1.3.** The **sample space** $S$ is the set of all elementary observation. Note, this could be finite or countably infinite.

**Definition 1.4.** An **event** is any subset of $S$.

**Definition 1.5.** The **intersection** of the set $A$ and $B$ is given by

$$A \cup B = \{E_i \in S : A \wedge E_i \in B\}.$$

**Definition 1.6.** The **union** of the set $A$ and $B$ is given by

$$A \cap B = \{E_i \in S : E_i \in A \vee E_i \in B\}.$$

**Definition 1.7.** The **complement** of the set $A$ is given by

$$\bar{A} = \{E_i \in S : E_i \notin A\}.$$

**Remark.** *Unions and intersection of events are still events i.e. subsets of $S$.*

## Axioms of probability

Given a sample space $S$, a probability $P(E)$ is assigned to every event $E \subset S$ such that:

1. $P(E) \geq 0$;
2. $P(S) = 1$;
3. If $A_1, A_2, \ldots, A_k$ is a collection of mutually exclusive events (i.e. if $A_i \cap A_j = \emptyset$ for $j \neq i$), then $P(A_1 \cup A_2 \cup \ldots \cup A_k) = \sum_{i=1}^{k} P(A_i)$. We also have the same result for an infinite sequence of events.

## 1.2 The Basic Principle of Counting

**Theorem 1.1.** With $m$ elements $a_1, a_2, \ldots, a_m$ and $n$ elements $b_1, b_2, \ldots, b_n$ one can exactly form $m \times n$ (ordered) pairs $(a_k, b_l)$.

*Proof.* Arrange the elements in an $m \times n$ array and count the number of elements in the array. $\square$

**Theorem 1.2.** With $n_1$ elements $a_1, a_2, \ldots, a_{n_1}; n_2$ elements $b_1, b_2, \ldots, b_{n_2}$ and $n_k$ elements $l_1, l_2, \ldots, l_{n_k}$, one can exactly form $n_1 \times n_2 \times \ldots \times n_k$ **multiplets** of $k$ elements.

**Definition 1.8.** Given $n$ different objects, the complete set of possible arrangements of these objects is called the set of **permutations.**

**Theorem 1.3.** The **number of permutations** of $n$ objects (all different and distinguishable) is given by

$$n! = n(n-1)(n-2)\ldots 1.$$

**Theorem 1.4.** The number of ordered arrangements of $r$ objects taken (without replacement) from a set of $n$ distinguishable objects is $P_r^n = \frac{n!}{(n-r)!} = n(n-1)\ldots(n-r+1)$.

**Definition 1.9.** A **combination** is a subset of $n$ objects taken from some larger set of $n$ distinguishable objects (where the order is irrelevant).

**Theorem 1.5.** The number of ways of selecting combinations of $r$ objects from a set of $n \geq r$ distinguishable objects is

$$C_r^n = \binom{n}{r} = \frac{n!}{(n-r)!r!}.$$

# 2 Some more combinatorics

**Definition 2.1.** A **partition** is an arrangements of $n$ elements into $k$ groups containing $n_1, n_2, \ldots, n_k$ objects respectively, with $0 \leq n_j \leq n$ and $n \sum_{j=1}^{k} n_j$. (Every element is assigned exactly to one group).

**Remark.** *The definition of partition is in the context of probability.*

**Theorem 2.1.** (Number of partitions). The number of partitions into $k$ distinct groups, containing $n_1, n_2, \ldots, n_k$ elements is

$$P_{n_1 n_2 \ldots n_k}^n = \binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \ldots n_k!}.$$

# 3 Probability rules

**Definition 3.1.** We say, $A_1, \ldots, A_k$ are **mutually exclusive** events if $A_i \cap A_j = \emptyset, i \neq j$.

**Distributive Law:**

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \text{ and}$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

**De Morgan's Law:**

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B} \text{ and}$$
$$\overline{(A \cap B)} = \overline{A} \cup \overline{B}.$$

**Theorem 3.1.** (Law of addition of probabilities)

$$P(A \cap B) = P(A) + P(A) - P(A \cup B).$$

# 4 Conditional probability

**Definition 4.1.** Probability of an event $A$ given an event $B$ has occurred is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

In the above definition, we should specify that $P(B) > 0$.

**Definition 4.2.** The event $A$ is **independent** of the event $B$ if $P(A|B) = P(A)$.

**Theorem 4.1.** The following statement are equivalent:

1. $P(A|B) = P(A)$, if $A$ is independent of $B$;
2. $P(B|A) = P(B)$, if $B$ is independent of $A$;
3. $P(A \cap B) = P(A)P(B)$.

**Definition 4.3. The Multiplicative Law of Probability:**
$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$, and if $A$ and $B$ are independent, $P(A \cap B) = P(A)P(B)$.

**Definition 4.4.** A family of sets $A_1, A_2, \ldots, A_n$ consisting of mutually exclusive and exhaustive events is called a **partition of the sample space** $S$. (Recall that this means that $A_i \cap A_j = \phi$ for $i \neq j$, and that $S = A_1 \cup A_2 \ldots \cup A_n$.

**Note:** Exhaustive events cover the probability of the whole sample space.

**Theorem 4.2. Properties of Independence**

1. If $A$ is independent of $B$, then $\overline{A}$ is independent of $B$.
2. If $A$ is independent of $B$, then $A$ is independent of $\overline{B}$.

**Theorem 4.3.** (Normalization of conditional probabilities.) If $A_1, A_2, \ldots, A_n$ is a partition of $S$, then $\sum_{i=1}^{n} P(A_i|B) = 1$.

# 5 Bayes' theorem

**Theorem 5.1.** Given 2 events $A$ and $B$, then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

**Theorem 5.2.** (Bayes' theorem, general version.) Given an event $A$ and the family $B_1, B_2, \ldots, B_n$ that is a partition of $S$,

$$\frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}.$$

# 6 Discrete random variables and their probability distributions

**Definition 6.1.** A **random variable** (sometimes r.v.) is a function defined on the sample space $S$ mapping each element of $S$ to a real number.

**Definition 6.2.** A **discrete random variable** is a random variable that can only take finitely or countably infinitely many distinct values. Note that, for any discrete random variable, we have the following two properties

1. $0 \leq p(x) \leq 1$;
2. $\sum_x p(x) = 1$.

## 6.1 Expectation value of a random variable

**Definition 6.3.** Given a random variable $X$ with probability distribution/probability mass function (p.m.f.) $p(x)$ its **expectation**, or expected value is defined as:

$$\sum_x xp(x).$$

The above sum is over all values $x$ that $X$ can take.

The expectation of any function $X$ is given by

$$E(f(x)) = \sum_x f(x)p(x).$$

Similarly, the expectation of $X^n$ is

$$E(X^n) = \sum_x x^n p(x).$$

**Definition 6.4.** The **variance** of $X$ is defined as

$$V(X) \equiv Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2.$$

**Theorem 6.1.** (**Properties of Expectations**). For any constants $a$ and $b$, and any function $f$ and $g$, we have:

1. $E(a) = a$;
2. $E(g(x)) = \sum_x g(x)p(x)$;
3. $E(af(X)) = aE(f(X))$;
4. (Linearity) $E(af(X) + bg(X)) = aE(f(x)) + bE(f(X))$.

**Theorem 6.2.** (**Properties of Variance**). For any constants $a$ we have:

1. $Var(a) = 0$;
2. $Var(X + a) = Var(X)$;
3. $Var(aX) = a^2Var(X)$;
4. $Var(aX \pm b) = a^2Var(X)$.

## 6.2 Standard discrete distributions

### 6.2.1 Binomial distribution

A random variable $X$ follows a binomial distribution if the following 5 conditions are satisfied:

1. There is a fixed number of trials ($n$).
2. Each trial results in either *'success'* or *'failure'*.
3. All the trials are independent.
4. The probability of 'success' ($p$) is the same in each trial.
5. The variable, $X$, is the total number of successes in the $n$ trials.

Then we can say that the probability mass function is $X \sim Bin(n, p)$, where $n$ is the number of trials and $p$ is the probability of the event being a *'success'*. The probability mass function is given by,

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n.$$

The expectation and variance are:

$$E(X) = np,$$
$$Var(X) = np(1-p).$$

### 6.2.2 Bernoulli distribution

A special case of the binomial distribution where $n = 1$. A discrete random variable which takes on only 2 values, typically $S = \{0, 1\}$.

$$p(x) = \begin{cases} p, & y = 1; \\ 1-p, & y = 0. \end{cases}$$

The expectations and variance are:

$$E(X) = x$$
$$Var(X) = p(1-p).$$

### 6.2.3 Geometric distribution

A random variable $X$ follows a geometric distribution if the following 3 conditions are satisfied:

1. There is a sequence of independent trials with only two possible outcomes – *'success'* and *'failure'*.
2. The probability of *'success'*,$p$, is constant.
3. $X$ is the number of trials until the first success occurs (including the *'successful'* trial itself).

Then the p.m.f. is $X \sim Geo(p)$, for $X = 1, 2, 3 \ldots$ with probability:

$$P(X = x) = p(1-p)^{x-1}.$$

The expectation and variance are:

$$\begin{aligned} E(X) &= \frac{1}{p} \\ Var(X) &= \frac{(1-p)}{p^2}. \end{aligned}$$

### 6.2.4 Hypergeometric distribution

The hypergeometric distribution concerns the case of sampling from a set containing a mixture of 2 different elements (red, black) without replacement. Suppose we have a set of size $N$ with $r$ red elements and $N-r$ black elements. Then the number of different ways of finding $x$ red elements in a draw of $n$ is $\binom{r}{x}\binom{N-r}{n-x}$. Hence the probability of finding x red elements is,for $x = 0, 1, \ldots, n$,

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}.$$

The expectation and variance are

$$\begin{aligned} E(X) &= \frac{nr}{N} \\ Var(X) &= \frac{nr(N-r)(N-n)}{N^2(N-1)}. \end{aligned}$$

#### 6.2.5 Poisson distribution

If $X$ represents the number of events that occur in a particular space or time, then $X$ will follow a Poisson distribution as long as:

1. The events occur randomly, and are independent of each other.
2. The events happen singly i.e. one at a time.
3. The events happen (on average) at a constant rate (either in space or time).

The Poisson parameter $\lambda$ is then the average rate at which these events occur (i.e. the average number of events in a given interval of space or time). Then the p.m.f. is $X \sim Po(\lambda)$, for $X = 0, 1, 2, \ldots$ with probability:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

The expectation and variance are:

$$\begin{aligned} E(X) &= \lambda \\ Var(X) &= \lambda. \end{aligned}$$

# 7 Moment generating functions

**Definition 7.1.** The $n^{\text{th}}$ moment of a random variable $X$ is given by

$$\mu_n = \sum_x x^n p(x).$$

In particular, $\mu_0 = 1$, and $\mu_1 = E(X)$. Recall that the variance of $X$ is given by

$$V(X) \equiv Var(X) = E(X^2) - (E(X))^2 = \mu_2 - \mu_1^2.$$

**Definition 7.2.** The **moment generating function** of $X$ is defined as

$$M_X(t) = E(e^{tX}) = \sum_x e^{tx} p(x).$$

**Theorem 7.1.** Moments of a distribution can be computed via derivatives of the moment generating functions.

$$\mu_k = E(X^k) = \frac{\mathrm{d}^k M_X(t)}{\mathrm{d}t^k}\Big|_{t=0}.$$

# 8 Continuous random variable (C.R.V.)

**Definition 8.1.** Let $X$ be a random variable whose set of possible values $S$ is a continuum of numbers such as an interval, i.e. the set of possible values is uncountable. We say that $X$ is a **continuous random variable** if there exists a non-negative function $f$ , such that the following property holds for any set $B$ of real numbers:

$$P(X \in B) = \int_B f(x)\,dx.$$

**Definition 8.2.** The function $f \geq 0$, piecewise continuous (i.e. continuous perhaps except for finitely many values of $x$) in the definition above is called the probability density function (p.d.f.) of the random variable $X$.

**Proposition 1.** The piecewise continuous function $f$, defined for every $x \in \mathbb{R}$ is a probability density function (p.d.f.) if

- $f(x) \geq 0$;
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$;
- $P(X = a) = \int_a^a f(x)\,dx = 0$.

**Definition 8.3.** Let $X$ be a c.r.v. with p.d.f. $f$. Then, the **cumulative distribution function** (c.d.f.) is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt.$$

Consequently, $F$ is continuous and the relationship between the p.d.f. and the cumulative distribution function $F$ is expressed by

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x} F(x).$$

**Theorem 8.1.** (**Properties of the C.D.F.**)

1. $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$;
2. $F(+\infty) = \lim_{x \to +\infty} F(x) = 1$;
3. $F(x)$ is a non decreasing function of $x$, i.e. if $x_1 < x_2$, then we have $F(x_1) \leq F(x_2)$.

**Remark.** *When drawing the c.d.f. you must draw the lines for which the function takes* 0 *and* 1.

## 8.1 Expectation values for a continuous random variable

**Definition 8.4.** The expectation value of a continuous random variable with p.d.f. $f(x)$ is

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx.$$

**Theorem 8.2.** (Properties of Expectations) For any constant $a$ and $b$, and any function $f$ and $g$,

1. $E(a) = a$;
2. $E(af(X)) = aE(f(X))$;
3. $E(af(X)+bg(X)) = aE(f(X))+bE(g(X))$, where $E(g(X)) = \int_{-\infty}^{\infty} g(x)p(x)\, dx$.

## 8.2 Quantiles

**Definition 8.5.** Given any random variable $X$, and $0 < p < 1$, the $p$-quantile of $X$, $\Phi_p$, is the smallest number such that

$$P(X \leq \Phi_p) = F(\Phi_p) \geq p.$$

## 8.3 Probability distribution of continuous random variable

### 8.3.1 The uniform random variable on an interval [a,b]

**Definition 8.6.** A random variable $X$ is **uniform** on the interval $[a, b]$, with $a < b$, if its p.d.f. is given by

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b.$$

[Notation: $X \sim U[a, b]$].

**Theorem 8.3.** If $X$ is a uniform random variable on the interval $[a, b]$ then

- $E(X) = \frac{a+b}{2}$;
- $Var(X) = \frac{(b-a)^2}{12}$;
- the c.d.f, $F$ is given by

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \begin{cases} 0, & y < a; \\ \frac{x-a}{b-a}, & a \leq x < b; \\ 1, & x \geq b. \end{cases}$$

### 8.3.2 The Gaussian or Normal distributed random variables

**Definition 8.7.** A random variable $X$ is said to be a **normal random variable**, or simply $X$ is normally distributed, with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ if the p.d.f. of $X$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

[Notation: $X \sim N(\mu, \sigma^2)$]. The p.d.f. is a bell-shaped curve that is symmetric about $\mu$.

**Theorem 8.4.** If $X \sim N(\mu, \sigma^2)$ then

- $E(X) = \mu$;
- $Var(X) = \sigma^2$;
- if $Y = aX + b$ then $Y \sim N(a\mu + b, a^2\sigma^2)$.

**Corollary 8.4.1.** *If $X \sim N(\mu, \sigma^2)$, then*

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

*Such a random variable is said to be a* ***standard normal random variable.*** *It is customary to denote the c.d.f. of $Z$ by $\Phi(x)$. That is,*

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$$

*The values of the standard c.d.f. for are given in Statistical tables.*

### 8.3.3 Normal approximation to the Binomial

**Theorem 8.5.** If $S_n$ denotes the number of successes that occur when $n$ independent trials, each resulting in a success with probability $p$, are performed, then, for any $x \in \mathbb{R}$,

$$P\left(\frac{S_n - np}{\sqrt{np(1-p)}}\right) \to_{n\to\infty} \Phi(x).$$

In general, this approximation is quite good when $np(1-p) \geq 10$.

### 8.3.4 Exponentially distributed random variable

**Definition 8.8.** A random variable $X$ whose probability density function is given, for some $\lambda > 0$, by

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0,$$

is said to be an **exponential random variable** with parameter $\lambda$ [notation: $X \sim Exp(\lambda)$].

**Remark.** *In practice, the exponential distribution often arises as the distribution of the amount of time until some specific event occurs.*

**Theorem 8.6.** If $X$ is an exponential random variable with parameter $\lambda$ then

- $E(X) = \frac{1}{\lambda}$;
- $Var(X) = \frac{1}{\lambda^2}$;
- the c.d.f., $F$ is given by

$$F(X) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

**Theorem 8.7.** The exponential distribution satisfies the **memoryless property.** That is, if $X \sim Exp(\lambda)$ then

$$P(X > t + s | X > t) = P(X > s),$$

for all $s, t > 0$.

**Proposition 2.** The time interval $T$ between two consecutive Poisson occurrences at the rate $\lambda > 0$ is exponentially distributed with mean value $\frac{1}{\lambda}$.

### 8.3.5 Γ-Distributed random variable

**Definition 8.9.** A random variable $X$ with p.d.f. given by

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0,$$

is said to be a **gamma distributed** random variable with parameters $(\alpha, \beta)$ [notation: $X \sim Gamma(\alpha, \beta)$].

The function $\Gamma(\alpha)$, called the **Gamma function**, is defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt.$$

**Proposition 3.**

$$\Gamma(n) = (n-1)!$$

#### 8.3.6 $\chi^2$ distribution

**Definition 8.10.** A random variable is said to have a $\chi^2$ distribution with $\nu$ degrees of freedom if it is a $\Gamma$-distributed random variable with $\alpha = \frac{\nu}{2}$ and$\beta = 2$, i.e.

$$Y \sim \chi^2_\nu \iff Y \sim \Gamma\left(\frac{\nu}{2}, 2\right).$$

# 9 Chebyshev's theorem

**Theorem 9.1.** Let $X$ be a random variable with finite mean $\mu$ and variance $\sigma^2$. Then, for any $k > 0$, we have:

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad \text{equivalently}$$
$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

# 10 Bivariate probability distributions

**Definition 10.1.** Let $(X, Y)$ be a random pair. We define the **joint cumulative distribution function** (joint c.d.f.) of $(X, Y)$ by

$$F(x, y) = P(X \leq x, Y \leq y), \quad (x, y) \in \mathbb{R}^2.$$

**Definition 10.2.** Let $(X, Y)$ be a random pair, where $X$ and $Y$ are both discrete random variables. We define the **joint probability mass function** (joint p.m.f.) of $(X, Y)$ by

$$p(x_i, y_j) = P(X = x_i, Y = y_j), \quad (x_i, y_j) \in X(\Omega) \times Y(\Omega)$$

The joint p.m.f. of $(X, Y)$ is usually given in the form of a table.

**Proposition 4.** The joint p.m.f. satisfies the conditions:

- $p(x_i, x_j) = P(X = x_i, Y = y_j) \geq 0, \quad \forall (x_i, x_j);$
- $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p(x_i, x_j) = P(X = x_i, Y = y_j) = 1.$

**Definition 10.3.** The functions

$$p_X(x_i) = \sum_{j=1}^{\infty} p(x_i, y_j) \text{ and}$$
$$p_Y(y_j) = \sum_{i=1}^{\infty} p(x_i, y_j)$$

are called the **marginal probability mass functions** of $X$ and $Y$, respectively.

**Definition 10.4.** The random variables $X$ and $Y$ are called jointly continuous if there exists a function $f(x, y)$ defined for $x, y \in \mathbb{R}$ such that, for every set $C$ of pairs of real numbers

$$P((X, Y) \in C) = \iint_{(x,y)\in C} f(x, y)\, dxdy.$$

The function $f(x, y)$ is called the **joint probability density function** of $X$ and $Y$.

**Proposition 5.** The joint p.d.f. satisfies the conditions:

- $f(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$;
- $\iint_{\mathbb{R}^2} f(x, y)\, dxdy = 1.$

**Definition 10.5.** The functions

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy \text{ and}$$
$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx$$

are called the **marginal probability density functions** of $X$ and $Y$, respectively.

## 10.1 Independent random variables

**Definition 10.6.** The random variables $X$ and $Y$ are said to be independent if, for any two sets of real numbers $A$ and $B$,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

In other words, $X$ and $Y$ are independent if the events $E = \{X \in A\}$ and $F = \{Y \in B\}$ are independent.

**Proposition 6.** The equality $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$holds if and only if, for all $x, y \in \mathbb{R}$,

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y).$$

In terms of the joint c.d.f. F of the random pair $(X, Y)$, $X$ and $Y$ are independent r.v.'s if

$$F(x, y) = F_X(x)F_y(y), \quad x, y \in \mathbb{R}.$$

## 10.2 Conditional distribution

**Definition 10.7.** Let $X$ and $Y$ be jointly discrete random variables. It is natural to define the **conditional probability mass function** of $X$ given $Y = y$ by

$$p_{X|Y}(x_i|y) = P(X = x_i|Y = y) = \frac{p(x_i, y)}{p_Y(y)},$$

for all values of $Y$ such that $p_Y(y) > 0$

# 11 DRAFT 2

# 12 Bivariate probability distributions

## 12.1 Joint distribution functions

**Definition 12.1.** The **joint cumulative distribution function** (or joint c.d.f.) of $X$ and $Y$, denoted by $F_{XY}(x, y)$, is the function defined by

$$F_{XY} = P(X \leq x, Y \leq y).$$

**Remark.** *The event $(X \leq x, Y \leq y)$ is equivalent to the event $A \cup B$, where $A$ and $B$ are events of $S$ defined by $P(A) = F_X(x)$ and $P(B) = F_Y(y)$.*

**Definition 12.2.** Two random variables $X$ and $Y$ are called **independent** if

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

for every value of $x$ and $y$.