Applied Probability for Computer Science

Isadora Antoniano-Villalobos
isadora.antoniano@unive.it

Master in Computer Science
Ca' Foscari University of Venice

Academic year 2023/2024

Overview of Random Variables and Distributions y

Equally likely outcomes

A simple situation for computing probabilities is the case of **equally likely outcomes**. That is, when the $\Omega = \{\omega_1, \ldots, \omega_n\}$, with each ω_i having the same probability.

$$\sum_{i=1}^{n} \mathbb{P}\left[\{\omega_i\}\right] = \mathbb{P}\left[\Omega\right] = 1 \quad \Rightarrow \quad \mathbb{P}\left[\{\omega_i\}\right] = 1/n.$$

And for any event $E \subset \Omega$,

$$\mathbb{P}\left[E\right] = \sum_{\omega \in E} \mathbb{P}\left[\{\omega\}\right] = \sum_{i=1}^{\#E} \frac{1}{n} = \frac{\#E}{\#\Omega}.$$

☞ **Combinatorics** help us count events in complex setups ➔ We will not cover this in the course, but it may be useful, make sure you know about **combinations and permutations, with and without replacement**

Random Variables

- Baron textbook **Chapter 3: Discrete Random Variables and Their Distributions** and **Chapter 4: Continuous Distributions** are part of the course prerequisites, but we will start with a quick overview to remember the main concepts

DEFINITION 3.1

A **random variable** is a function of an outcome,

$$X = f(\omega).$$

In other words, it is a quantity that depends on chance.

- The domain of a random variable is the sample space Ω
- The range, also called the **support** of the random variable is a subset of $\mathbb{R}$, corresponding to the possible values the random variable can potentially take. Interesting cases are: $\mathbb{R}$, $\mathbb{N}$, $(0, \infty)$, $(0, 1)$

Discrete Random Variables and their Distributions

- If a random variable can only take a finite or at most countable number of values, we call it **discrete**

DEFINITION 3.2

Collection of all the probabilities related to X is the **distribution** of X. The function

$$P(x) = \boldsymbol{P}\{X = x\}$$

is the **probability mass function**, or **pmf**. The **cumulative distribution function**, or **cdf** is defined as

$$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{y \leq x} \boldsymbol{P}(y). \tag{3.1}$$

The set of possible values of X is called the **support** of the distribution F.

➔ **Example 3.1** Let X be the number of heads when tossing 3 fair coins.

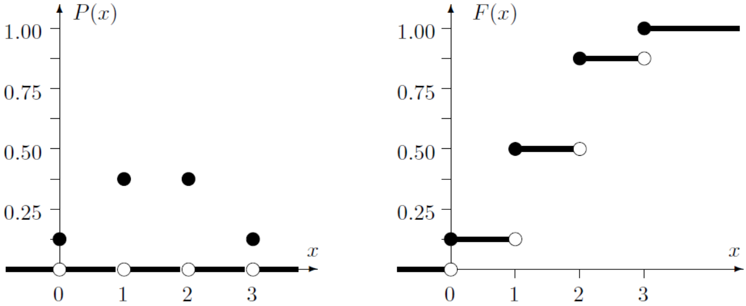


FIGURE 3.1: *The probability mass function $P(x)$ and the cumulative distribution function $F(x)$ for Example 3.1. White circles denote excluded points.*

Discrete Random Variables and their Distributions

- Once an experiment is completed, and the outcome $\omega \in \Omega$ is known, the random variable X takes one and only one value $X(\omega) = x$, therefore, the collection of events $\{[X = x] : x \in \mathbb{R}\}$ form a partition and:

$$\sum_x P(x) = \sum_x \mathbb{P}\left[X = x\right] = 1$$

  - In particular, for any event A,

$$\mathbb{P}\left[X \in A\right] = \mathbb{P}\left[A\right] = \sum_{x \in A} P(x)$$

- Notice that, if $P(x) > 0$ for an uncountable number of values X, this sum would go to ∞. This is why a special treatment is needed for random variables with uncountable support

Continuous Random Variables and their Distributions

DEFINITION 4.1

Probability density function (pdf, density) is the derivative of the cdf, $f(x) = F'(x)$. The distribution is called **continuous** if it has a density.

☞ Formally, a random variable is called a **continuous random variable** if it has a continuous cdf, $F(x) = \mathbb{P}[X \leq x]$. There is a subtle difference, but in this course we won't worry about it

- For continuous random variables, the support is uncountable and the probability mass function is always equal to zero:

$$P(x) = 0 \quad \text{for all } x$$

Continuous Random Variables and their Distributions

- The pdf plays a role analogous to the pmf. Indeed, $f(x) \geq 0$ for all x and

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

- In particular, for any event A

$$\mathbb{P}[X \in A] = \mathbb{P}[A] = \int_A f(x)dx$$

Example 4.1. The lifetime, in years, of some electronic component is a continuous random variable with the density

$$f(x) = \begin{cases} \dfrac{k}{x^3} & \text{for} \quad x \geq 1 \\ 0 & \text{for} \quad x < 1. \end{cases}$$

Find k, draw a graph of the cdf $F(x)$, and compute the probability for the lifetime to exceed 5 years.

Properties of Probability Distributions

Distribution	Discrete	Continuous
Definition	$P(x) = P\{X = x\}$ (pmf)	$f(x) = F'(x)$ (pdf)
Computing probabilities	$\boldsymbol{P}\{X \in A\} = \sum_{x \in A} P(x)$	$\boldsymbol{P}\{X \in A\} = \int_A f(x)dx$
Cumulative distribution function	$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{y \leq x} P(y)$	$F(x) = \boldsymbol{P}\{X \leq x\} = \int_{-\infty}^{x} f(y)dy$
Total probability	$\sum_x P(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$

- In both cases, the cdf $F(x)$ is a **non-decreasing** function of x, taking values in the interval $[0, 1]$,with

$$\lim_{x \downarrow -\infty} F(x) = 0 \qquad \lim_{x \uparrow \infty} F(x) = 1$$

Distributions of Random Vectors

- For simplicity, we focus on a vector (X, Y) of dimension 2, but everything can be extended to a vector $X_1, \dots X_n$ of dimension n.
- The **joint probability mass function** or **joint pdf** of X and Y is

$$P(x, y) = \mathbb{P}\left[(X, Y) = (x, y)\right] = \mathbb{P}\left[X = x \cap Y = y\right]$$

- In this course, we focus on vectors in which components are all discrete or all continuous

DEFINITION 4.2

For a vector of random variables, the **joint cumulative distribution function** is defined as

$$F_{(X,Y)}(x, y) = \boldsymbol{P}\left\{X \leq x \ \cap Y \leq y\right\}.$$

The **joint density** is the *mixed derivative* of the joint cdf,

$$f_{(X,Y)}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x, y).$$

Distributions of Random Vectors

Distribution	Discrete	Continuous
Marginal distributions	$P(x) = \sum_y P(x,y)$ $P(y) = \sum_x P(x,y)$	$f(x) = \int f(x,y)dy$ $f(y) = \int f(x,y)dx$
Independence	$P(x,y) = P(x)P(y)$	$f(x,y) = f(x)f(y)$
Computing probabilities	$\boldsymbol{P}\{(X,Y) \in A\}$ $= \sum_{(x,y)\in A} \sum P(x,y)$	$\boldsymbol{P}\{(X,Y) \in A\}$ $= \iint_{(x,y)\in A} f(x,y)\,dx\,dy$

Distributions of Random Vectors

DEFINITION 3.4

Random variables X and Y are **independent** if

$$P_{(X,Y)}(x, y) = P_X(x)P_Y(y)$$

for *all* values of x and y. This means, events $\{X = x\}$ and $\{Y = y\}$ are independent for all x and y; in other words, variables X and Y take their values independently of each other.

☞ If the random variables are continuous, we must use the pdf instead of the pmf to check independence. We need to verify

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{for all } (x, y)$$

Distributions of Random Vectors

Example 3.6. A program consists of two modules. The number of errors, X, in the first module and the number of errors, Y, in the second module have the joint distribution, $P(0,0) = P(0,1) = P(1,0) = 0.2$, $P(1,1) = P(1,2) = P(1,3) = 0.1$, $P(0,2) = P(0,3) = 0.05$. Find (a) the marginal distributions of X and Y, (b) the probability of no errors in the first module, and (c) the distribution of the total number of errors in the program. Also, (d) find out if errors in the two modules occur independently.

Some important properties of a distribution

☞ Baron textbook **Chapter 3: Discrete Random Variables and Their Distributions** and **Chapter 4: Continuous Distributions** are part of the course prerequisites, but we will start with a quick overview to remember the main concepts

1. Expectation

DEFINITION 3.5

> **Expectation** or **expected value** of a random variable X is its mean, the average value.

- Since X can take different values with different probabilities, its average value is not just the average of all its values. Rather, it is a **weighted average**

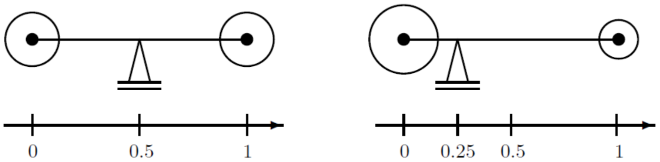
Some important properties of a distribution

Expectation, discrete case

$$\boxed{\mu = \mathbf{E}(X) = \sum_x xP(x)}$$

- **Examples 3.7 and 3.8:** Imagine X can only take two values, 0 and 1
  - **(a)** $P(0) = P(1) = 0.5$
  - **(b)** $P(0) = 0.75; \quad P(1) = 0.25$
- The expectation can be interpreted as the **center of mass** or **center of gravity** of the distribution

(a) $\mathbf{E}(X) = 0.5$ (b) $\mathbf{E}(X) = 0.25$

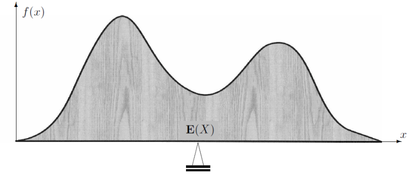


Some important properties of a distribution

- As usual, for continues random variables, in order to calculate the expectation, we use the pdf and an integral in place of the pmf and the sum respectively:

$$\mu = \mathbb{E}\left[X\right] = \int x f(x) dx$$

But the interpretation as the center of gravity of the distribution still holds



Some important properties of a distribution

- The **expectation**, also called **mean** or **expected value** is a property of the distribution, so random variables with the same distribution will also have the same expectation

  Example: Consider a single toss of a fair coin and define 2 variables, X takes the value 0 for tails and 1 otherwise. Y takes the value 1 for tails and 0 otherwise. Clearly, X and Y are different variables (if one takes the value 0, the other takes the value 1), but they have the same distribution and therefore the same mean

- The **expectation of a function** of a random variable X, say $Y = g(X)$ can be calculated from the distribution of X. Indeed, depending on whether the variables are discrete or continuous, we have:

$$\mathbb{E}\left[g(X)\right] = \sum_x g(x)P(x) \quad \text{or} \quad \mathbb{E}\left[g(X)\right] = \int g(x)f(x)dx$$

Some important properties of a distribution

Properties of expectations

$$\mathbf{E}(aX + bY + c) = a\,\mathbf{E}(X) + b\,\mathbf{E}(Y) + c$$

In particular,

$$\begin{aligned} \mathbf{E}(X+Y) &= \mathbf{E}(X) + \mathbf{E}(Y) \\ \mathbf{E}(aX) &= a\,\mathbf{E}(X) \\ \mathbf{E}(c) &= c \end{aligned}$$

For **independent** X and Y,

$$\mathbf{E}(XY) = \mathbf{E}(X)\,\mathbf{E}(Y)$$

2. Variance and Standard Deviation

DEFINITION 3.6

Variance of a random variable is defined as the expected squared deviation from the mean. For discrete random variables, variance is

$$\sigma^2 = \mathrm{Var}(X) = \mathbf{E}\,(X - \mathbf{E}X)^2 = \sum_x (x-\mu)^2 P(x)$$

- The variance can also be computed as

$$\mathsf{Var}\,[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}\,[X])^2$$

DEFINITION 3.7

Standard deviation is a square root of variance,

$$\sigma = \mathrm{Std}(X) = \sqrt{\mathrm{Var}(X)}$$

3. Covariance and Correlation

DEFINITION 3.8

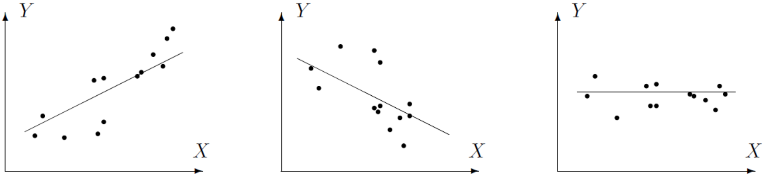
Covariance $\sigma_{XY} = \mathrm{Cov}(X, Y)$ is defined as

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbf{E}\{(X - \mathbf{E}X)(Y - \mathbf{E}Y)\} \\ &= \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y) \end{aligned}$$

It summarizes interrelation of two random variables.

- Two random variables X and Y can be **positively correlated**, **negatively correlated** or **uncorrelated** depending on whether the covariance is positive, negative or zero

Some important properties of a distribution



(a) $\mathrm{Cov}(X,Y) > 0$ (b) $\mathrm{Cov}(X,Y) < 0$ (c) $\mathrm{Cov}(X,Y) = 0$

- If X and Y are independent, then $\mathrm{Cov}[X,Y] = 0$, but the converse is not necessarily true: two uncorrelated variables may be dependent

DEFINITION 3.9

Correlation coefficient between variables X and Y is defined as

$$\rho = \frac{\mathrm{Cov}(X,Y)}{(\mathrm{Std}X)(\mathrm{Std}Y)}$$

Some important properties of a distribution

- $|\rho| = 1$ happens only when all the possible values of the pair (X, Y) lie on a straight line. In this case, we say that they are **perfectly correlated** (positively or negatively)

Some important properties of a distribution

Properties of variances and covariances

$$\mathrm{Var}(aX + bY + c) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y)$$

$$\begin{aligned}&\mathrm{Cov}(aX + bY, cZ + dW)\\ &\quad = ac\,\mathrm{Cov}(X,Z) + ad\,\mathrm{Cov}(X,W) + bc\,\mathrm{Cov}(Y,Z) + bd\,\mathrm{Cov}(Y,W)\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{Cov}(Y,X)\\ \rho(X,Y) &= \rho(Y,X)\end{aligned}$$

In particular,

$$\begin{aligned}\mathrm{Var}(aX + b) &= a^2\,\mathrm{Var}(X)\\ \mathrm{Cov}(aX + b, cY + d) &= ac\,\mathrm{Cov}(X,Y)\\ \rho(aX + b, cY + d) &= \rho(X,Y)\end{aligned}$$

For independent X and Y,

$$\begin{aligned}\mathrm{Cov}(X,Y) &= 0\\ \mathrm{Var}(X + Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y)\end{aligned}$$

Some important properties of a distribution

Discrete	Continuous
$\mathbf{E}(X) = \sum_x xP(x)$	$\mathbf{E}(X) = \int x f(x) dx$
$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \sum_x (x-\mu)^2 P(x)$ $= \sum_x x^2 P(x) - \mu^2$	$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \int (x-\mu)^2 f(x) dx$ $= \int x^2 f(x) dx - \mu^2$
$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \sum_x \sum_y (x-\mu_X)(y-\mu_Y)P(x,y)$ $= \sum_x \sum_y (xy)P(x,y) - \mu_x\mu_y$	$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \iint (x-\mu_X)(y-\mu_Y) f(x,y)\, dx\, dy$ $= \iint (xy) f(x,y)\, dx\, dy - \mu_x\mu_y$

<u>NOTATION</u>

μ or $\mathbf{E}(X)$	=	expectation
σ_X^2 or $\mathrm{Var}(X)$	=	variance
σ_X or $\mathrm{Std}(X)$	=	standard deviation
σ_{XY} or $\mathrm{Cov}(X,Y)$	=	covariance
ρ_{XY}	=	correlation coefficient

4. Chebyshev's Inequality

Chebyshev's inequality

$$P\{|X-\mu| > \varepsilon\} \leq \left(\frac{\sigma}{\varepsilon}\right)^2$$

for any distribution with expectation μ and variance σ^2 and for any positive ε.

- Knowing just the expectation and variance, one can find the range of values most likely taken by this variable. Since X belongs to the interval $[\mu-\varepsilon, \mu+\varepsilon]$ with probability of at least $1-(\sigma/\varepsilon)^2$

☞ Chebyshev's inequality shows that only a large variance may allow a variable X to differ significantly from its expectation μ. In this case, the risk of seeing an extremely low or extremely high value of X increases. For this reason, risk is often measured in terms of a variance or standard deviation.