## Contents



# Part IV

# Continuous Random Variables

## 13 Introduction

**Review: Random Variable**

Recall the definition of a random variable.

<u>Definition</u>: *random variable*

> Given a random experiment with sample space $S$ consisting of simple events $\omega$, a random variable $X$ is a function that maps $\omega \in S$ to the real numbers $\Re$.
>
> $$X : \omega \to \Re$$

The *range* $X(S)$ of the random variable is the collection of numbers in $\Re$ that the random variable maps to.

A *discrete random variable* is one where $X(S)$ is countable, i.e. there is a one-to-one map between the elements of $X(S)$ and the positive integers. In other words, one can imagine labeling each simple event $\omega \in X(S)$ with a *unique* integer.

Now we consider cases where $X(S)$ is uncountable, i.e. there are too many numbers to label each one with a unique integer. Any line segment, e.g. $[a, b)$, $[a, b]$, $(a, b]$, $[a, \infty)$, $(a, \infty)$, $(-\infty, b]$, or $(-\infty, b)$, for finite $a, b \in \Re$, is uncountable (Real analysis covers these ideas in detail.). We are talking about *continuous random variables*, although we will be more precise with our definition later.

Examples of such random variables are

- The time until the bus arrives at the bus stop. $X(S) = [0, \infty)$.
- The probability a person will die from heart disease, each probability depends on a person's genes, life experience, exposures, etc. $X(S) = [0, 1]$.
- The distance a discus lands from the discus thrower. $X(S) = [0, \infty)$.
- The distance a dart lands from the bulls-eye given that it does not hit the bulls-eye but hits the dart board. $X(s) = (r, R]$, where $r$ is the radius of the bulls-eye and $R$ is the radius of the board.

## 13.1 Probability Mass Function Does Not Exist

Previously, we argued that the probability mass function (pmf) $p_X(x) = P(X = x)$ completely defines a discrete random variable. Everything to know about a discrete random variable is summarized in the pmf.

So, we would like to define a pmf for these new random variables that exist on intervals. Unfortunately, it is not possible. We will show this by contradiction.

Suppose we could define $P(X = x)$ for all $x \in X(S)$. Then, it must be true (by the Axioms of Probability) that

$$\sum_{x \in X(S)} P(X = x) = 1 \tag{3}$$

Consider the subset of $X(S)$ whose probability exceeds $\frac{1}{n}$:

$$A_n = \left\{ x : P(X = x) > \frac{1}{n} \right\}$$

Then, the sum over just this subset is

$$\sum_{x \in A_n} P(X = x) > \sum_{x \in A_n} \frac{1}{n} = \frac{|A_n|}{n}$$

where $|A_n|$ is the number of elements in $A_n$. The above "sub-sum" only converges if $A_n$ is finite: $|A_n| < \infty$. Thus, for all $n$, $|A_n| < \infty$ in order for eq 3 to converge. Furthermore, the range $X(S)$ is a countable union of $A_n$:

$$X(S) = A_2 \cup A_3 \cup \cdots = \cup_{n=2}^{\infty} A_n$$

A countable union of finite sets is countable. Thus, we have reached a contradiction. Either $X(S)$ is countable, in which case $X$ is a discrete random variable, or $P(X = x) = 0$ for all but a countable subset of $X(S)$, which again makes $X$ finite.

The main conclusions are the pmf does not exist and $P(X = x) = 0$.

## 13.2 Probability Distribution

### 13.2.1 Cumulative Density Function (cdf)

We'll have to come up a novel way to describe random variables of this type.

Definition: *cumulative distribution function (cdf)*

For any random variable $Y$, the cumulative distribution exists is defined as

$$F(y) = P(Y \leq y), \quad -\infty < y < \infty$$

Notice, the cdf is defined for all possible values of $y \in \Re$.

Properties: cdf

1. The cdf is defined for discrete and continuous random variables.
2. $\lim_{y \to -\infty} F(y) = \lim_{y \to -\infty} P(Y \leq y) = 1 - P(Y \in Y(S)) = 0$.
3. $\lim_{y \to -\infty} F(y) = \lim_{y \to -\infty} P(Y \leq y) = P(Y \in Y(S)) = 1$.
4. For $y_1 < y_2$, $F(y_1) \leq F(y_2)$, so $F(y)$ is a monotonic, non-decreasing function of $y$.
5. For $y_1 < y_2$, $P(y_1 < Y \leq y_2) = P(Y \leq y_2) - P(Y \leq y_1) = F(y_2) - F(y_1)$.

Definition: *continuous random variable*

A continuous random variable is one for which $F(y)$ is continuous (and with derivative defined at all but a finite number of points).

### 13.2.2 Probability Density Function (pdf)

Definition: *probability density function (pdf)*

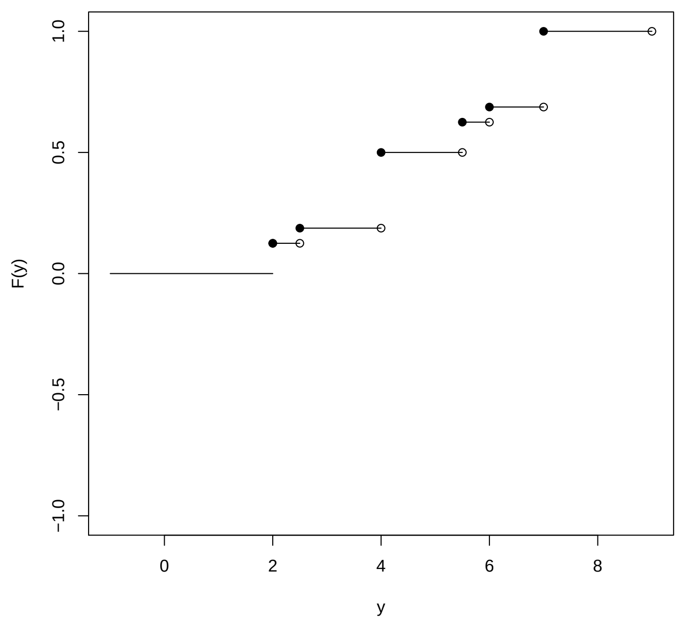
For a continuous random variable, the probability density function if it exists is

$$f(y) = \frac{dF(y)}{dy}$$

Properties: pdf

1. When the pdf exists, $F(y) = \int_{-\infty}^{y} f(t)dt$
2. $f(y) \geq 0$ because $F(y)$ is non-decreasing
3. The pdf is *not* a probability, i.e. it is possible for $f(y) > 1$.
4. $\int_{-\infty}^{\infty} f(t)dt = 1$ because $F(y) \to 1$ as $y \to \infty$.

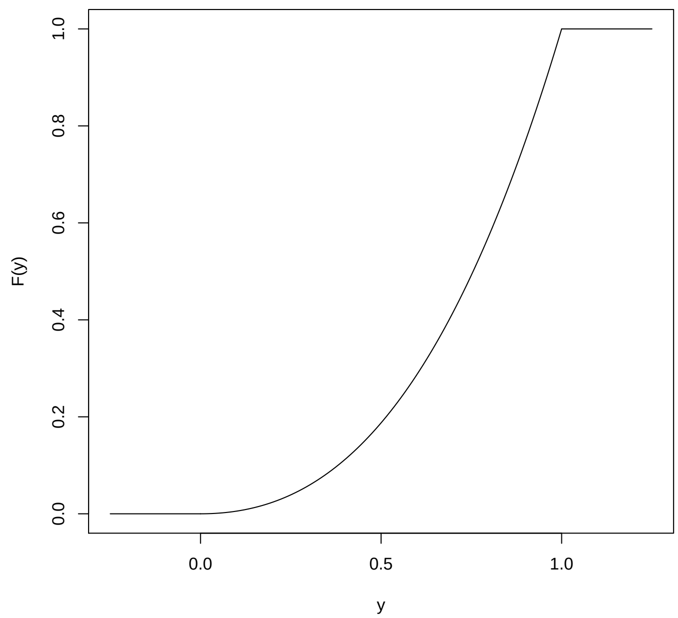
### 13.2.3 Examples




An example of the cdf of a discrete random variable (in graphical form and as formula).

$$F(y) = P(Y \leq y) = \begin{cases} 0, & \text{for } y < 2 \\ 1/8, & \text{for } 2 \leq y < 2.5 \\ 3/16, & \text{for } 2.5 \leq y < 4 \\ 1/2, & \text{for } 4 \leq y < 5.5 \\ 5/8, & \text{for } 5.5 \leq y < 6 \\ 11/16, & \text{for } 6 \leq y < 7 \\ 1, & \text{for } y \geq 7 \end{cases}$$

An example of a continuous random variable cdf (graphical and formula).

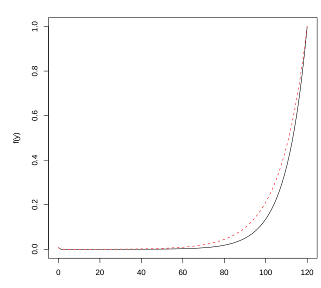



$$F(y) = \begin{cases} \frac{y^3}{2} + \frac{y^2}{2}, & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

What is the pdf?

$$f(y) = \frac{3y^2}{2} + y, \quad 0 \le y \le 1$$

**Example: continuous random variable**




Suppose you know that waist-to-hip ratio affects life expectancy. In particular, if $Y$ is life expectancy and $w$ is waist-to-hip ratio, then the pdf of life expectacy is given by (made up)

$$f(y) = C\left[0.0075e^{-100y} + e^{-\frac{0.07(y-120)}{w}}\right], \quad y \ge 0$$

where $C$ is chosen such that $f(y)$ is a pdf.

Questions that you should be able to answer:

- If $w = 0.7$, what is the value of $C$?

- What is the cdf $F(y)$?
- What is $P(Y > 65 + c)$ for some constant $c$, i.e. that someone lives $c$ years past retirement.
- What is $P(50 \leq Y \leq 70)$?
- What is $P(Y > 80 \mid Y > 65)$?