Chapter 1

Introduction

1.1 Recalling basics of randomness

Recalling the concept of randomness
What is a random variable?
What can we predict about a random variable? Connecting to the distribution of its probability.

The concept of randomness is all-pervasiveness in our daily lives. Let us review the basic concepts related to randomness, by considering the following example.

- Timmy wants to embark upon a new fiscal routine of investing the monthly disposable income within his household, into the stock market, where "disposable income" is defined as the difference between the money that is reaching the household every month from Timmy's and his partner's salaries, and the money that is outgoing in monthly expenditure. However, he cannot for sure know what this amount of "disposable income" is going to be, at the end of a month in the future, since this amount will vary depending on:
(*) expenses that the household will face in that month - that is not known to Timmy apriori;
(*) variations in income that might affect Timmy and his partner in that considered month.
Unless the incoming and outgoing amounts during that considered month in the future, are known, the exact value of "disposable income" remains an unknown. Usually, the incoming and outgoing amounts relevant to a month into the future will be known only after that month has passed.

So "disposable income" is a variable, and the value that it takes for any trial of the "experiment", is unknown, prior to the undertaking of a trial of this experiment.
So

"disposable income" is a random variable,

where the

"experiment" is: record monthly outgoings and incomes relevant to the household.

Let us recast all this English into succinct statements using notation.
–Monthly disposable income is denoted by X;
–monthly income into the household is denoted by X_i;
–monthly expenditure in this household is denoted by X_o.
–Then $\boxed{X := X_i - X_o.}$
–Experiment: record X_o and X_i every month from 1-st to n-th month.
We said that X is a random variable (abbreviated to r.v. from now on).
Here X_i and X_o are also r.v.s (same motivation that permits calling X an r.v.
–As this experiment is one such that (abbreviated as s.t. from now on), **we do not know the outcome of a trial of this experiment for sure, unless this trial has been conducted, we call this experiment a "probabilistic experiment".**
–Any other kind of experiment in which we can predict the outcome of a trial for sure, even before this trial of the experiment is conducted is called a "deterministic experiment".
–Here, r.v. X_o is a real-valued r.v., s.t. it is known to never be negative. So $X_o \in \mathbb{R}_{\geq 0}$.
–Timmy and his partner are sure that they will never lose their jobs, and that their wages will be paid every month. So X_i is positive-definite: $X_i \in \mathbb{R}_{>0}$.
–$X = X_i - X_o$ can be positive or negative. So $X \in \mathbb{R}$.
–Also, let outcome of the p-th trial of this experiment defined above - which is equivalent to the p-th of the n months that the outgoings and incomes are tracked - be $X_o = x_o^{(p)}$ and $X_i = x_i^{(p)}$. Note that here $p = 1, \ldots, n$. Then in this p-th month, $X = x_p = x_i^{(p)} - x_o^{(p)}$, $\forall p = 1, \ldots, n$.
–*Notation alert:*
(*) value taken by a real-valued r.v. called "A", will be designated "a" from now on;
(*) $\forall$ stands for "for all";
(*) space that a r.v. lives in, will be identified often.

–So we see that conducting any trial of the experiment allows for recording values of relevant r.v.s; here, the p-th trial of the probabilistic experiment allows for the recording of values of r.v.s X_i and X_o in the p-th month. **So some r.v.s are s.t. their values can be known by undertaking an experiment; the value of any such r.v. noted in a given trial of the experiment, changes in general, to another value, in another trial of the same experiment. Values of such observed r.v.s in the p-th trial of the experiment, then inform on the "outcome" of the experiment in this p-th trial.**
–Let us refer to the outcome ω_p from the p-th trial of this experiment to be denoted as the vector $\omega_p = (x_i^{(p)}, x_o^{(p)})^T$, where $X_o = x_o^{(p)}$ and $X_i = x_i^{(p)}$ in this p-th trial of the experiment, $\forall p = 1, \ldots, n$.
–We also see that whatever the outcome is, in any trial of a conducted experiment, this outcome can be used to define a new r.v. Thus, value of r.v. X is defined using outcome of a trial of the undertaken probabilistic experiment. In our simple example, value of r.v. X is simply defined using the 2 components of vector ω_p.
–In general, a r.v. is defined as some kind of a function of the outcome of a probabilistic experiment.
–We denote this intuition mathematically by stating: r.v. $X = f(\Omega)$, where $f(\cdot)$ is a function that takes the outcome as its input, and outputs the r.v.

r.v. are functions of outcomes of a probabilistic experiments.

–By definition, unless the p-th trial of the probabilistic experiment is undertaken, we do not know the value that r.v. X takes in this p-th instance. But, having observed monthly household income and expenditures in the past, Timmy knows that he could assign (rough rankings to) probability of the event that X takes the value x_p. For instance, Timmy knows that probability for x_p to be $\pounds 1000$ is "low", while that for x_p to lie in the interval $-\pounds 100$ to $\pounds 100$ is "comparatively higher". Let us cast these in notation:

$$\Pr(X = 1000) \text{ is low.}$$

$$\Pr(X \in [-100, 100]) \text{ is higher, where monies are in GBP.}$$

Suppose Timmy can do better than simply assign a rank to the probability of an event that $X = x_p$. In other words, he has an objective way of quantifying this probability, and this allows him to compute

$$\Pr(X = 1000) = 0.1.$$

$$\Pr(X \in [-100, 100]) = 0.7 \text{ where monies are in GBP.}$$

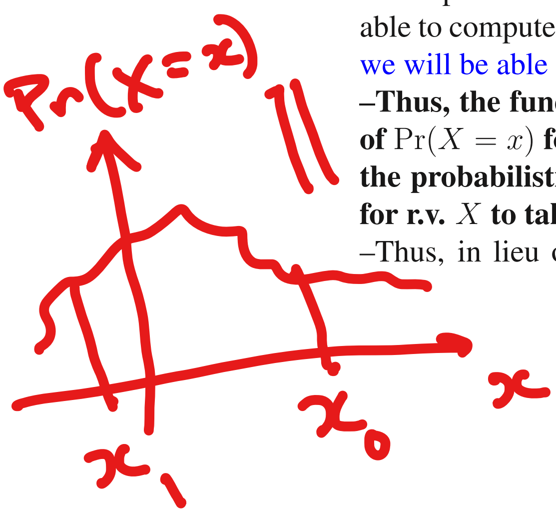
–In fact, let Timmy's "objective probability calculator" allow him to state that his household affairs are in such a state that the maximal probability for the event $X = x_p$ is attained when $x_p = \pounds 12$.
–In addition, Timmy knows that as x_p deviates further from $\pounds 12$, probability for the event $X = x_p$ declines in a way that is captured with the algebraic form:

$$\Pr(X = x_p) = K \exp\left(-\frac{(x_p - 12)^2}{0.39^2}\right),$$

where the constant $K = 1/\sqrt{2\pi \times .39^2}$.
–Timmy's partner thinks that the variation in probability of the event that $X = x_p$, with deviation of x_p away from $\pounds 12$ is again symmetric about $\pounds 12$, but best described by the algebraic form:

$$\Pr(X = x_p) = \frac{1}{(x_p - 12)^2 + a^2},$$

where a is the value of the parameter A in this model for probability of the event $X = x_p$, that Timmy's partner does not know the value of. So A remains an unknown model parameter in this model for the variation in the probability [of the event that monthly disposable income in their household attains value x_p], with x_p.
–Thus, what Timmy and is partner have advanced are forms of such variation in the probability that r.v. X attains a value, with this value. i.e. variation in $\Pr(X = x_p)$, with x_p.
–We do not need to call this value "x_p" necessarily, but we could call it by any other notation. To ease notational burden, let us just call it x; there is no loss of information in this easing of notation, as we generalise beyond our simple experiment.
–So what Timmy and his partner have advanced, are two different models for the variation in $\Pr(X = x)$, with x.
–If all parameters of any such model of $\Pr(X = x)$ are known to us, we will be able to compute the value of the probability that r.v. X takes a value x, $\forall x$. Then we will be able to predict the probability, with which the r.v. X takes a value x.
–Thus, the fundamental advantage of knowing the model for the variation of $\Pr(X = x)$ for ant r.v. X, is that we can predict - even before any trial of the probabilistic experiment has been conducted - what the probability is, for r.v. X to take the value x.
–Thus, in lieu of knowing such a value x that r.v. X will attain for sure, at



the end of any trial of the experiment, we have establish a means of knowing the probability for such an event to occur, before the probabilistic experiment is conducted.
–So we do not know that $X = x$ for sure, but know that this event will occur with a certain probability. In other words, we know about the value that is taken by r.v. X, less than surely $\rightsquigarrow$ we know of the value that X will attain, with uncertainty $\rightsquigarrow$ we know the event that $X = x$, with a probability < 1 in general.
–Thus, values attained by random variables are known probabilistically, i.e. with uncertainty.
–As long as we know what the probability is with which r.v. X attains value x, we can quantify the uncertainty in our knowledge of the value that r.v. X attains. Such quantification of uncertainty of our knowledge of the state of a r.v. is possible as long as we know the model for variation of $\Pr(X = x)$, with x.
–We will soon explore details of such a model that allows us to know of how $\Pr(X = x)$ varies with x. –Before we embark upon that study, let us pause and remind ourselves, that in the discussions above, we did not consider discussions of what probability is - let alone define it. Also, our interest in identifying the space that any generic mathematical structure lives in, has not been extended to identify host spaces of (*) outcomes; (*) events; (*) probabilities. We will undertake all these discussions soon.
–One last theoretical digression: if r.v. X is a function of the outcome ω, why do you not seek $\Pr(X(\omega) = x)$, and instead try to compute $\Pr(X = x)$. This is simply an abbreviation.

1.2 Continuous random variables

$X \in \mathbb{R}$

The correct question to ask.
Probability Density Function & Cumulative Distribution Function

- The question:

  ”what is the probability with which X attains value x?”

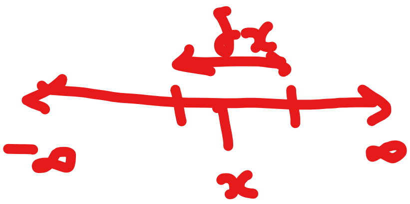
  is meaningless, and will not be used again - unless qualified - given that disposable income $X \in \mathbb{R}$.
  –Why does the residence of real-valued r.v. X in the real line, i.e. in $\mathbb{R}$, render the question “what is $\Pr(X = x)$?” wrong?

as that "what is $\Pr(X=x)$" is wrong, because one cannot assign an identifiable value x to a real X.

–Well, X is in $\mathbb{R}$.
(*) Assume:

we are able to identify an isolated, arbitrarily chosen point x on the real-line.

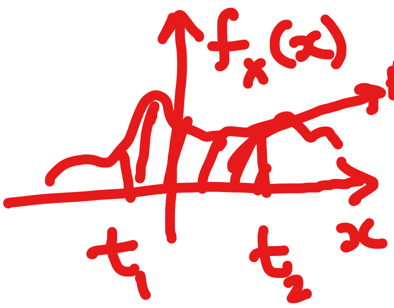
(*) Then its "neighbourhood", of arbitrary width $\delta x > 0$, will be "**dense**", i.e. between any pair of points in this neighbourhood, there are infinite other real numbers.
(*) As x and $\delta x > 0$ are arbitrary, denseness of the δx-neighbourhood of x holds $\forall x \in \mathbb{R}$ and $\forall \delta x > 0$, no matter how small δx is, including when δx approaches 0, i.e. when $\delta x \longrightarrow 0$, i.e. when $x + \delta x \longrightarrow x$.
(*) Thus, even when $\delta x \longrightarrow 0 \equiv x + \delta x \longrightarrow x$, the δx-wide neighbourhood of x is populated with infinite neighbours. In other words, even then identifiability of x on the real line no longer holds. Therefore, x cannot be identified as an isolated point on the real line.
(*) Therefore initial assumption is contradicted, i.e. an arbitrarily chosen point x cannot be identified on the real line.



(*) Hence we cannot ask the question "what is $\Pr(X = x)$?".
(*) We will amend this question to ask "what is $\Pr(X \in [x - \delta x/2, x + \delta x/2])$?" where we denote the width of the "immediate neighbourhood" of x to be δx.
–To be clear, **let us define "immediate neighbourhood" of x - of width δx - to be the (infinitesimally) narrow neighbourhood around the number x, s.t. probability for X to take a value anywhere inside that neighbourhood, does not change as we move from one point inside that neighbourhood to another point inside it.**

As $\delta x \to 0$
$\Pr(X \in$ [narrow interval]$)$
is a constant

–I want to avoid making the cumbersome computation of $\Pr(X \in [x - \delta x/2, x + \delta x/2])$, and yet, want to know how the probability for r.v. X to attain values in a given interval, varies with the location of this interval in the real line. How do I do this?
–The solution is to resort to a function - that we shall call a "density" - where this density function of the r.v. X is computed at x, as below:



$$\text{AREA} = \Pr(X \in [t_1, t_2]) = \int\limits_{u=t_1}^{u=t_2} f_X(u) du, \text{ where } t_1 \leq t_2;\ t_1, t_2 \in \mathbb{R}.$$

–We update intuition on the "smallness" of the δx defined above, to state that δx is small s.t. the density $f_x(x)$ is the same at all $x \in [x - \delta x/2, x + \delta x/2]$.
–Then in the definition of $f_X(\cdot)$ above, for t_1 set to be equal to $x - \delta x/2$ and $t_2 = x + \delta x/2$, $f_X(x)$ being a constant within the interval $[x - \delta x/2, x + \delta x/2]$,

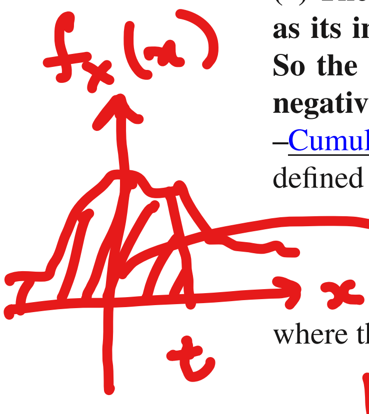
the right hand side of the last equation reduces to $f_X(x)(x+\delta x/2-(x-\delta x/2)) = f_X(x)\delta x$,so that

$$\Pr(X \in [x - \delta x/2, x + \delta x/2]) = f_X(x)\delta x.$$

Then as width δx approaches 0, the last equation allows us to define the density function as

$$f_X(x) = \underset{\delta x \longrightarrow 0}{limit} \frac{\Pr(X \in [x - \delta x/2, x + \delta x/2])}{x + \delta x/2 - (x - \delta x/2)} \equiv \underset{\delta x \longrightarrow 0}{limit} \frac{\Pr(X \in [x, x + \delta x])}{x + \delta x - (x)} \equiv \frac{d\Pr(x)}{dx},$$

where we have recalled from the definition of derivative.
–Thus, the density $f_X(\cdot)$ readily informs on how the distribution of the probability that X takes values in a given interval.
–Remember that
(*) $0 \leq \Pr(X = x \in [t_1, t_2]) \leq 1$**,** $\forall t_1, t_2 \in \mathbb{R}; t_1 \leq t_2$**.**
(*) The probability density function $f_X(\cdot)$ **is a function that takes a value** x **as its input, and outputs the slope of** $\Pr(X = x \in [x - dx/2, x + dx/2])$**. So the probability density can take values anywhere in the space of non-negative reals, i.e.** $f : \mathbb{R} \longrightarrow \mathbb{R}_{\geq 0}$**.**
–Cumulative Distribution Function or cdf of r.v. X, computed at any value t is defined as



AREA :=

$$F_X(t) := \Pr(X \in (-\infty, t]) \equiv \int_{-\infty}^{t} f_X(x)dx,$$

$\Longrightarrow f_x(x) \geq 0$

where the Probability Density Function or pdf of r.v. X is

$$f_X(x) := \frac{dF_X(x)}{dx}.$$

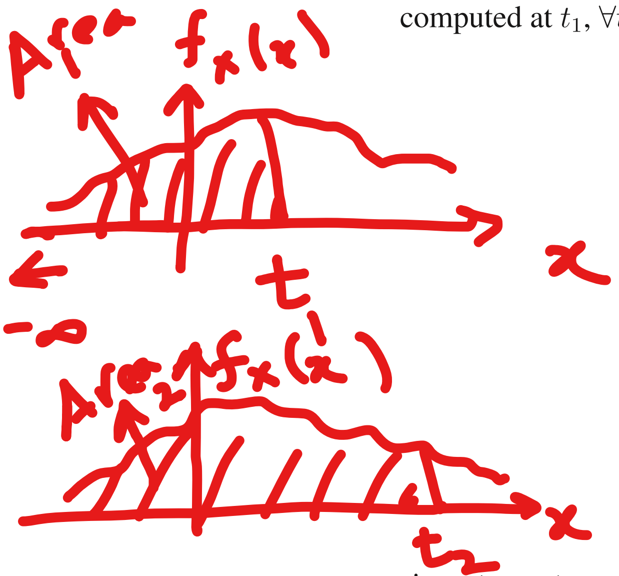
Here we consider the smallest value that is attainable by r.v. X, to be s.t. it approaches -∞; this is contained in our earlier declaration that r.v. X lives in the space of reals, i.e. $X \in \mathbb{R}$.

- Then seeking knowledge of variation of the probability for r.v. $X \in \mathbb{R}$ to attain a value that lies in the interval $[t_1, t_2] \forall t_1 \leq t_2; t_1, t_2 \in \mathbb{R}$, with the location of (the centroid of) this interval in $\mathbb{R}$, is equivalent to seeking the pdf of X at any x, i.e. $f_X(x)$. Then we can compute

$$\Pr(X = x \in [t_1, t_2]) = F_X(t_2) - F_X(t_1) = \int_{-\infty}^{t_2} f_X(x)dx - \int_{-\infty}^{t_1} f_X(x)dx.$$

Alternatively, we could know the cdf $F_X(x)$, and then can compute the pdf $f_X(x)$ as the derivative of the cdf with respect to x. Using the pdf we can compute $\Pr(X = x \in [t_1, t_2])$.

- Notice that the cdf computed at $t_2 > t_1$, is always greater than or equal to that computed at t_1, $\forall t_2, t_1 \in \mathbb{R}$, i.e.



$$F_X(t_2) \geq F_X(t_1), \text{ for } t_2 > t_1, \text{ since}$$

$$F_X(t_2) = \int_{-\infty}^{t_2} f_X(x)dx, \text{ while}$$

$$F_X(t_1) = \int_{-\infty}^{t_1} f_X(x)dx \implies$$

$$F_X(t_2) - F_X(t_1) = \int_{t_1}^{t_2} f_X(x)dx \geq 0,$$

since $t_1 < t_2$ and pdf $f_X : \mathbb{R} \longrightarrow \mathbb{R}_{\geq 0}$. Thus, the cdf is a monotonic non-decreasing function.

Then: Area$_2 \geq$ Area$_1$ $\forall t_2 \geq t_1$

- Additionally, a real-value r.v. surely attains a value that lies somewhere in the space of the reals, i.e.

$$\Pr(X = x \in (-\infty, \infty)) = 1, \text{ or}$$

$$F_X(\infty) = 1, \text{ or,}$$

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

1.3 Discrete random variables

String-valued variables
Discrete r.v.s and Probability Mass Function