The Hitchhiker's Guide
to
the Risk-Neutral Galaxy

Nicolae Santean, PhD, PDF, CQF

4th Edition

Every morning in the wild, a gazelle awakens. One thing is sure for the gazelle that day, as every other… She must run faster than the fastest lion. If she cannot, she will be killed and eaten.

Every morning, a lion awakens. For the lion too, one thing is certain… This day and every day, he must run faster than the slowest gazelle.

Whether fate names you a gazelle or a lion, is of no consequence. It is enough to know that with the rising of the sun, you must run, and you must run faster than the day before, for the rest of your days, or you will perish.

African tale

Contents

1 Prologue

The initial reason for writing this essay was to provide the background and a derivation of the IR quanto adjustment, as a mean of expressing the Hull-White dynamics of foreign currency interest rates under a domestic risk-neutral measure. However, one thing led to another, and the essay ended up including three different applications of the change of measure (or, Girsanov Theorem); namely, quanto adjustment, stock price dynamics under the risk-neutral measure, and pricing IR derivatives under the T-forward measure. The presented approach is neither meant to be original nor mathematically rigorous: it just aims at building an intuition of the employed methodology, hence facilitating a quick understanding of change of measure technique and some of its important applications. It is structured as a “user manual” and targets an audience with little background on the matter. The essay is organized in three main parts, as follows:

In the first part we provide an overview of the basic notions and notations used across the essay. We introduce a few elementary concepts of measure theory and emphasize that a probability function is a particular instance of a measure. Random variables and stochastic processes are added to the picture, and along with them we introduce the fundamental notion of filtration as an information carrier and main driver of conditional probabilities and conditional process expectations. Perhaps the single most important concept of this chapter is that of martingale; that is, a stochastic process expressing the dynamics of a fair game. Brownian motions (BM) are built next, from scratch: starting with a sequence of coin tosses, turned into a random walk, which is then scaled to a discrete approximation of a standard BM, and finally, culminating with a geometric BM. Some important properties of BM, such as Gaussian, martingale and fractal properties, are highlighted.

Two important results are closing this introductory chapter: Itô’s Lemma – the building brick of stochastic calculus, and Doléans-Dade exponential – the solution of one of the most elementary stochastic differential equations (SDE). Perhaps more important than the formulae themselves, are the notions of quadratic variation and covariation, used all across the essay. They reflect the fundamental difference between deterministic and stochastic functions.

The second part of the essay lays the foundation of the change of measure technique, as given by Girsanov Theorem. We do not present the theorem in its most general form, rather just enough details needed for the applications that follow. The theorem is presented in two flavors: for random variables, and for stochastic processes. The presentation is informal, emphasizing the intuition at the expense of rigorousness. A few examples are provided, to allow the reader build some mental images around this technique. The simulation example in particular, provides a concrete instance of Girsanov Theorem in action. This section culminates with a first application of the change of measure technique in the derivation of so-called quanto adjustment for foreign interest rates.

Perhaps the most important application of change of measure in practice is the formulation of derivative pricing models. The third part of this essay introduces three probability measures: physical (real-world), risk-neutral, and T-forward measures, and uses these measures in pricing of a few simple derivatives.

2 Part A: Developing a Toolbox

"The beginning of wisdom is to call things by their right names."
Confucius, 551 – 479 BC

2.1.1 Probability Spaces

A *sample space* Ω is a set of possible *outcomes* of a stochastic experiment. We call the elements of Ω simply "outcomes". By 2^{Ω} we denote the set of all subsets of Ω. Any element of 2^{Ω} is an *event*, that is, a collection of possible outcomes of the experiment. We say that the event $A \in 2^{\Omega}$ has been *realized* during an experiment if the outcome of the experiment belongs to A. We sometime say that the event A has *occurred* during the experiment.

A σ*-algebra* $\mathcal{F}$ is a *subset* of 2^{Ω}, that satisfies the following conditions:

1. (non-emptiness) $\mathcal{F}$ is non-empty.
2. (complement closure) if $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$.
3. (closure over countable unions) if $A_1, A_2, ..., A_n, ... \in \mathcal{F}$ then $A_1 \cup A_2 \cup ... \cup A_n \cup ... \in \mathcal{F}$.

Notes:

- $\Omega, \varnothing \in \mathcal{F}$: if $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$, $\Omega = \mathrm{A} \cup (\Omega \setminus A) \in \mathcal{F}$, and $\varnothing = \Omega \setminus \Omega \in \mathcal{F}$.
- The smallest/coarsest σ-algebra on Ω is $\mathcal{F}_0 = \{\varnothing, \Omega\}$, and the largest/finest is $\mathcal{F}_{\infty} = 2^{\Omega}$.

A *sub-*σ*-algebra* of $\mathcal{F}$ is any σ-algebra $\mathcal{G}$ included in $\mathcal{F}$: $\mathcal{G} \subseteq \mathcal{F}$, $\mathcal{G}$ is a σ*-algebra* over Ω.

A <u>*measurable space*</u> is the ordered pair $(\Omega, \mathcal{F})$. A *measurable set* is any set $A \in \mathcal{F}$, and A is often called an $\mathcal{F}$*-measurable set. If* $\mathcal{G}$ is a *sub-*σ*-algebra* of $\mathcal{F}$, any $\mathcal{G}$*-measurable set* is an $\mathcal{F}$-measurable set, but not necessarily vice versa.

A *measure* on $(\Omega, \mathcal{F})$ is any function $m : \mathcal{F} \to \mathbb{R}$ that satisfies the following properties:

1. (positive) $m(A) \geq 0, \quad \forall A \in \mathcal{F}$
2. (null empty set) $m(\varnothing) = 0$
3. (countable additive) if $\{A_i\}_{i \in I}$ is a countable collection of disjoints sets in $\mathcal{F}$, then $m\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} m(A_i)$.

A <u>*measure space*</u> is an ordered tuple $(\Omega, \mathcal{F}, m)$ with m being a well-defined measure on $(\Omega, \mathcal{F})$ (note the terminology: measure space vs. measurable space).

A *probability measure* is any measure $P : \mathcal{F} \to [0,1]$, that satisfies $P(\Omega) = 1$.

A *probability space* is an order tuple $(\Omega, \mathcal{F}, P)$, with P being a well-defined probability measure. It is a measure space where the measure is a probability measure. Any element $\omega \in \Omega$ is an outcome of a given stochastic experiment, and any set of outcomes $\mathrm{A} \subseteq \Omega$ is an event that, if it belongs to $\mathcal{F}$, is assigned a chance of occurrence (that is, a probability): $P(A)$.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(\Gamma, \mathcal{T})$ be a measurable space. A *random variable* is any function $X : \Omega \to \Gamma$ that satisfies $E \in \mathcal{T} \Rightarrow f^{-1}(E) \in \mathcal{F}$. In other words, the pre-image of any measurable set in Γ is a measurable set in Ω (with respect to their σ*-algebras,* respectively).

A *metric space* is a pair $(\mathcal{M}, d)$ where $\mathcal{M}$ is a non-empty set and $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ (the *distance* on $\mathcal{M}$) is a function that satisfies the following conditions, for any $x, y, z \in \mathcal{M}$:

1. (positivity) $\quad d(x,y) \geq 0$
2. (identity of indiscernibles) $\quad d(x,y) = 0 \Leftrightarrow x = y$
3. (symmetry) $\quad d(x,y) = d(y,x)$
4. (triangle inequality) $\quad d(x,z) \leq d(x,y) + d(y,z)$

An *open set* in $\mathcal{M}$ is any set $U \subseteq \mathcal{M}$ that satisfies

$$\forall x \in U,\ \exists \varepsilon \in \mathbb{R}^{+} \text{ such that } \forall y \in \mathcal{M} \text{ with } d(x,y) < \varepsilon \ \Rightarrow\ y \in U,$$

that is, once an element x belongs to U, an entire "disk" centered in x is included in U.

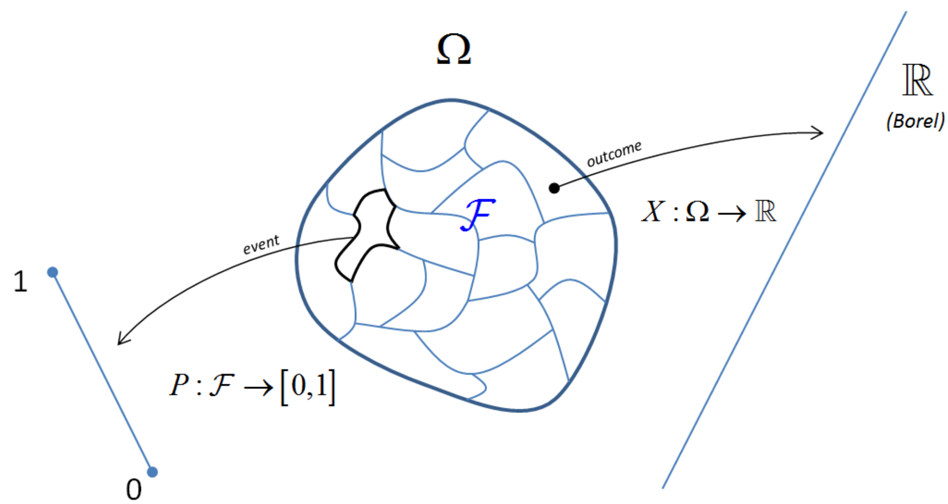
A *closed set* is the complement of some open set.

A *Borel* σ*-algebra* is the smallest σ-algebra containing all closed sets of $\mathcal{M}$.

Notable example. The *Borel* σ*-algebra on the real numbers* ($\mathcal{M} = \mathbb{R}$) is generated by the collection of closed sets $\{\ (-\infty, r]\ |\ r \in \mathbb{R}\ \}$. By "generated", we mean performing a σ-closure on a collection of sets: adding all combinations of countable unions, complements, and the empty set, to the collection. It is the smallest σ-algebra containing the initial collection.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(\mathbb{R}, d)$ be the Borel σ-algebra on the real numbers with the usual distance given by the interval length ($d(a,b) = b - a$). A *real-valued random variable* is any function $X:\Omega\to\mathbb{R}$ that satisfies $\forall r\in\mathbb{R}:\ \left\{\ \omega\ \middle|\ X(\omega)\le r\right\}\in\mathcal{F}$. Consequently, the pre-image of any Borel set (that is, Borel-measurable set) is an $\mathcal{F}$-measurable set (that is, a measurable set in Ω with respect to $\mathcal{F}$). Formally, we write $\mathcal{F}\supseteq\left\{X^{-1}(B)\middle|B\text{ is a Borel set}\right\}$. Terminology-wise, we sometimes say that *X is an $\mathcal{F}$-measurable random variable*. Note carefully that the definition doesn't make use of probability, hence we could very well start with "Let $(\Omega,\mathcal{F})$ be a measurable space and…". However, the term "random" in "random variable" reminds us that the definition becomes more meaningful in the context of a probability space indeed. Note also that the collection of pre-images of all Borel sets $\left\{X^{-1}(B)\middle|B\text{ is a Borel set}\right\}$ is itself a σ-algebra, more precisely, a sub-σ-algebra of $\mathcal{F}$. We call it the σ-algebra induced by the random variable X.

The following figure puts together the notions defined so far (note that $\mathcal{F}$ need not be a partition, and is not so indeed – we depicted it this way only for aesthetical reasons):

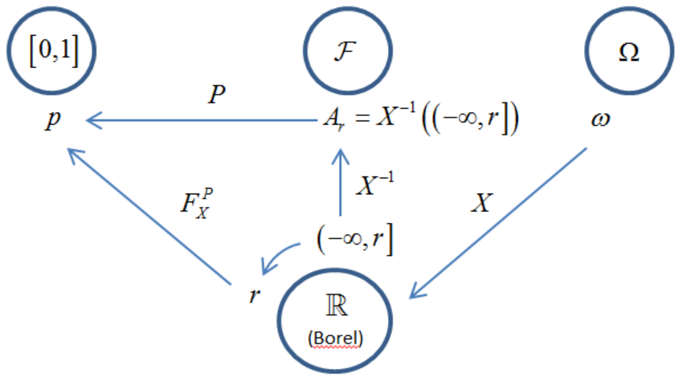


In the following, we assume the fundamental notion of Lebesgue integral known. Let $(\Omega,\mathcal{F},P)$ be a probability space and X be a real-valued random variable, $X:\Omega\to\mathbb{R}$. The *cumulative distribution function* (CDF) of X is the function $F_X^P:\mathbb{R}\to[0,1]$ given by the expression $F_X^P(r)=P(X\le r)$. This function is well-defined indeed:

$$event(X\le r)=X^{-1}\left((-\infty,r]\right)=\left\{\ \omega\in\Omega\ \middle|\ X(\omega)\le r\right\}\in\mathcal{F}.$$

The *probability density function* (PDF) of X is the function $f_X^P:\mathbb{R}\to\mathbb{R}$ given by $f_X^P(x)=\dfrac{dF_X^P(x)}{dx}$; thus, f_X^P is the instantaneous rate of change of F_X^P. In other words, $f_X^P(x)$ quantifies the likelihood that X will materialize in the neighborhood of x, that is, in $(x-dx,\ x+dx)$. Note that unlike F_X^P,

f_X^P does not denote a probability (can be greater than 1): is rather a concentration of probability. It is sometimes called the *distribution* of X. This is a sketch for the construction of F_X^P :



The (*unconditional*) expectation of X under probability P is given by $E^P[X] = \int_\Omega X(\omega) dP(\omega)$. This is, intuitively, the probability-weighted average of X over Ω (since $P(\Omega)=1$). For some set $A \subset \Omega$, denote by 1_A the identity over A and null everywhere else. The (*conditional*) expectation of X over A is $E^P[X \mid \omega \in A] = E^P[1_A X] = \int_A X(\omega) dP(\omega)$, which can intuitively be viewed as a weighted sum of the values of X over A, sum weighted by the probability weights $dP(\omega)$. It gives the expected value of X given (conditioned by) the realization of the event A. Note carefully that here (and almost everywhere, for that matter) we use Lebesgue integrals, to allow integration over peculiar domains.

Under the above context, let $\mathcal{G} \subset \mathcal{F}$ be a sub-sigma algebra of $\mathcal{F}$. The *conditional expectation* $E^P\left[X \middle| \mathcal{G}\right]$ (that is, *expectation conditional on a* *filter* $\mathcal{G}$) is a $\mathcal{G}$-measurable random variable which satisfies $\int_A E^P\left[X \middle| \mathcal{G}\right](\omega) dP(\omega) = \int_A X(\omega) dP(\omega)$ for all $A \in \mathcal{G} \subseteq \mathcal{F}$. Informally, $E^P\left[X \middle| \mathcal{G}\right]$ has the same expectations as X over all events in $\mathcal{G}$. Note carefully that the first Lebesgue integral is over $\mathcal{G}$-measurable sets and the second one is over $\mathcal{F}$-measurable sets. Note also that, if X happens to be $\mathcal{G}$-measurable as well, then $E^P\left[X \middle| \mathcal{G}\right] = X$.

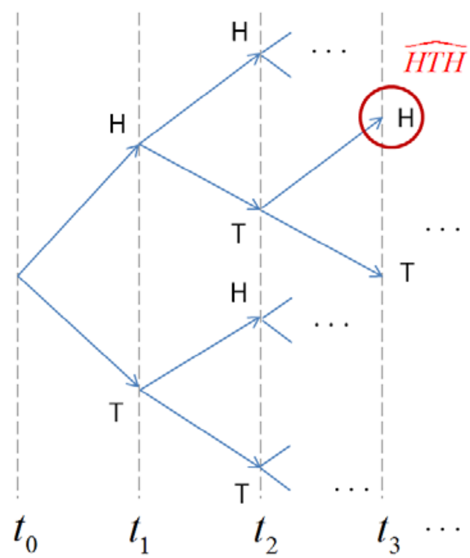
A *stochastic process* is a time-indexed set of real-valued random variables $\{X_t\}_{t \in T}$ (note that we usually consider an index taking values from a countable set; yet here, the index is from a most-likely uncountable, yet ordered, subset of $\mathbb{R}$: some period of time). Considering a process as the stochastic evolution of some system in time, the random variable X_t represents the state of the system as observed at time t, and is more accurately written as $X_t(\omega)$, to reflect the dependence on an outcome in Ω. We may imagine every time instance being associated a random variable (hence a stochastic experiment) that materializes (and takes a random value) at the given time. Here $\omega \in \Omega$ can be viewed as a sequence of random elementary outcomes $\omega = o_1 o_2 ... o_t ...$, with o_i occurring at time t_i (o_i is the realization of X_i, at time t_i). Note carefully the difference between the elementary outcomes o_i occurring at an instant time t_i (and governing the random variables X_i) and the outcomes $\omega \in \Omega$, representing a random *evolution* of the system (ω is a sequences of random experiments which produce a time-indexed sequence of random variable values). The difference will become apparent shortly. Usually, the random variables are iid (independent and identically distributed), as they all are reflections of some idiosyncratic properties of a same system.

2.1.2 Filtrations

"I don't know why we are here, but I'm pretty sure that it is not in order to enjoy ourselves."
Ludwig Wittgenstein, 1889 – 1951

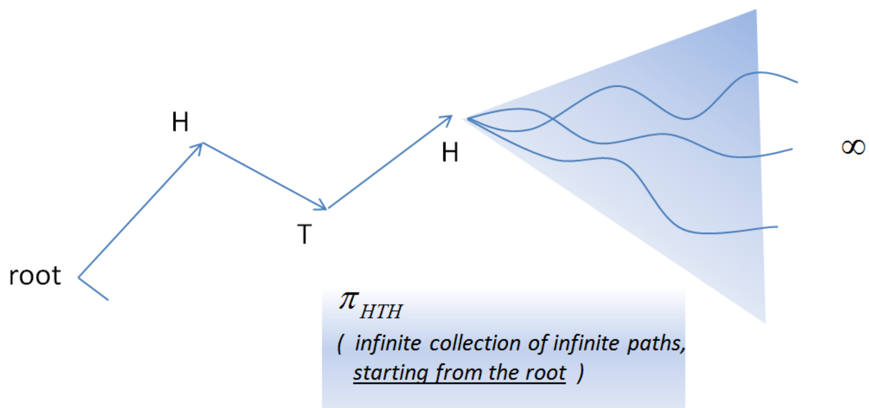
Yet again, let Ω be a set of outcomes. Consider Γ an ordered collection of time-points $t_0 = 0, t_1, t_2, ...$ (note that, despite the generic notation, Γ may not be countable, yet must necessarily be an ordered set), and consider a collection of σ-algebras $\{\mathcal{F}_t\}_{t\in\Gamma}$ over Ω that satisfies: $\mathcal{F}_0 = \{\varnothing, \Omega\}$, and for any $t_i < t_j$, $\mathcal{F}_{t_i} \subset \mathcal{F}_{t_j}$. Denote by $\mathcal{F}_\infty$ the σ-closure of the union of all these σ-algebras (since one is included into the other, one can potentially write $\mathcal{F}_\infty = \lim_{t\to\infty}(\mathcal{F}_t)$). That is, $\mathcal{F}_\infty$ is the *smallest σ-algebra* that contains all $\mathcal{F}_t$'s. Further, equip $\mathcal{F}_\infty$ with a probability measure P so that $(\Omega, \mathcal{F}_\infty, P)$ is structured as a probability space. The collection $\{\mathcal{F}_t\}_{t\in\Gamma}$ is called a *filtration* on $(\Omega, \mathcal{F}_\infty, P)$. Note that $(\Omega, \mathcal{F}_t, P/\mathcal{F}_t)$ is a probability *subspace* of $(\Omega, \mathcal{F}_\infty, P)$, where by $P/\mathcal{F}_t$ we understand the restriction of P to the domain $\mathcal{F}_t$. In the following, we will use the term "filtration" to refer to either $\{\mathcal{F}_t\}_{t\in T}$ or any individual $\mathcal{F}_t$, or $\mathcal{F}_\infty$, without creating confusion. Sometimes we refer to a particular σ–algebra of a filtration as a "filter". To summarize, is useful to view $(\Omega, \mathcal{F}_t, P/\mathcal{F}_t)$ as a probability space induced by the filtration $\mathcal{F}_t$. We have just constructed a time-indexed collection of probability spaces $\{(\Omega, \mathcal{F}_t, P/\mathcal{F}_t)\}_{t\in\Gamma}$. Finally, $\mathcal{F}_\infty$ is sometimes simply denoted by $\mathcal{F}$, by dropping the subscript. Here we introduced a "global" probability measure, which can be restricted to any filtration $\mathcal{F}_t$; yet, in a more general sense, each $(\Omega, \mathcal{F}_t)$ measurable space can have its own associated probability P_t - we just simplify the framework, for a better intuition.

Example of filtration. Consider a process governed/generated by a coin toss: a toss at time t_i is represented by the random variable X_i. More precisely, we toss a coin at successive points in time, denoted by t_i, and each outcome o_i will evolve the system from a previous state into a new state, with the transition governed by some pre-defined rules. Here the coin represents a stochastic factor/perturbation/entropy that governs the transition of the system from one state to the other. The actual random outcomes, from the system evolution point of view, are described as $\omega = o_1 o_2 \ldots o_t \ldots$, consisting of a successive sequence of coin toss outcomes: Ω consists of these possible evolutions ω. The entire evolution of this process/system can be described by a binary tree (a discrete case), with each node representing a possible state of the system at a given time (H and T represent heads and tails, respectively). The infinite branches spanning from the root of the tree represent the outcomes in Ω, that is, an evolution of the system in cause. Here's a depiction:



We aim at conferring a probability structure to this binary tree, hence we start by defining some tree terminology. The notions of root, node, edge and path are considered understood. In order to label paths, consider the binary alphabet $\Sigma = \{H, T\}$, and a word $w \in \Sigma^*$ (here, by "word" we understand a sequence of symbols/letters, empty, finite or infinite, and Σ^* denotes the collection of all such words). We can use these "words" to denote paths (partial or infinite) in the tree, which start from origin. Indeed, say, the word $w = HTH$ will label the path starting from origin, visiting the node H, then the node T, and arriving at another node H. From now on, we will make no distinction between a path and the word that labels it. For convenience, when w is finite, we denote by $\widehat{w}$ the terminal node. If λ denotes the "empty word" (a word with no symbols), $\widehat{\lambda}$ will denote the tree's root. In our example, $\widehat{HTH}$ denotes the node marked in red in the above figure, reached from the root and following the path $w = HTH$. This notation doesn't apply to infinite words/paths, for obvious reasons.

Of our main interest are infinite paths and sets/collections of such paths. Notation wise, by π_w we denote the set of all infinite paths that start from the root, and their labels start with w. For illustration, if $w = HTH$, then π_{HTH} represents the set of all paths that start from the root, traverse the nodes $\widehat{H}$, $\widehat{HT}$ and $\widehat{HTH}$, and then continue traversing other nodes, infinitely. In language notation, one can view $\pi_w = w*$, that is, the set of all infinite words/paths that have w as a prefix. Visually:



Note carefully that $\pi_u \supset \pi_{uv}$, for any labels u and v. Indeed, if one finite path u is the prefix of another finite path uv, any infinite path in π_{uv} (that is, which starts with uv) must necessarily be in π_u, as it obviously starts with u as well.

Define $\Omega = \pi_\lambda$, that is, Ω is the set of all possible infinite paths starting from the root. By the above observation, $\pi_w \subset \Omega$, for any w, as the empty word is the trivial prefix of any word.

Back to our process, the set of possible outcomes of one coin toss is $\{H,T\}$: with the already-introduced elementary outcome notation, $o_i \in \{H,T\}$ is a coin toss outcome at time t_i, and one can see Ω as the set of all possible evolution paths $\omega = o_1 o_2 ... o_t ...$ of the system in time. In this analogy, the infinite path/word ω is an evolution of the process. Note carefully the difference between the coin toss elementary outcomes, and the process' evolution. In this terminology, the set π_w (with w finite path) represents the set of all process evolution paths that all start the same: they start by traversing w. In probability terms, π_w is a *process event* (a collection of process evolutions).

Now, let's define filtrations. At time t_0, we have the filter $\mathcal{F}_0 = \{\varnothing, \Omega\}$. It is a filter that contains only two events: the impossible and the assured event. Information-wise, this is equivalent to a state of total ignorance, meaning that we have no information about the process' evolution (present or future). This filter consists of two trivial events: the null event $\varnothing$, with associated probability $P(\varnothing) = 0$, and the universal event Ω, with associated probability $P(\Omega)=1$. This conveys the idea that the system *"will certainly evolve in time", i.e.* $P(\varnothing)=0$, and *"anything is possible", i.e.* $P(\Omega)=1$.

At time t_1 we have $\mathcal{F}_1$ obtained from $\mathcal{F}_0$ by considering the coin toss outcomes $o_1^1=H$ and $o_1^2=T$, which lead to the possible system states at time t_1: $w_1^1=H$ and $w_1^2=T$, which further lead to all possible events anticipated at time t_1: π_H and π_T. Here, π_H is the set of all possible evolutions of the system, given that it has started with H, and π_T is defined similarly. But since $\mathcal{F}_1$ must be a σ-algebra, it should also include all possible combinations (countable unions and complements) of these events, as well as the events in $\mathcal{F}_0$ (in order to have $\mathcal{F}_0\subset\mathcal{F}_1$). Thus, $\mathcal{F}_1=\{\varnothing,\Omega,\pi_H,\pi_T\}$, since $\pi_H\cup\pi_T=\Omega$ (any path in Ω must start with either H, the case in which it is in π_H, or with T, the case in which it is in π_T), then $\pi_H\setminus\pi_T=\pi_H$ (as the events are independent), and $\overline{\pi_H}=\pi_T$ (we already stated that $\pi_H\cup\pi_T=\Omega$), etc. . Here we used the notation $\overline{A}$ to denote the complement of A ($\overline{A}=\Omega\setminus A$).

The probabilities associated with the new events in $\mathcal{F}_1$ are obviously $P(\pi_H)=P(\pi_T)=0.5$ (if we use a fair coin) and we have already defined the probabilities of the previous events in $\mathcal{F}_0$. This fully defines $(\Omega,\mathcal{F}_1,P/\mathcal{F}_1)$.

At time t_2 we follow the same steps. We obtain $\mathcal{F}_2$ by first adding to $\mathcal{F}_1$ new events: π_{HH}, π_{HT}, π_{TH}, and π_{TT}, and then performing a σ-closure of what we have so far, that is, we add all possible combinations of these events and the events in $\mathcal{F}_1$. Thus,

$$\mathcal{F}_2=\{\varnothing,\Omega,\pi_H,\pi_T,\pi_{HH},\pi_{HT},\pi_{TH},\pi_{TT},\ldots,\overline{\pi_{HH}},\overline{\pi_{HT}},\ldots,\pi_{HH}\cup\pi_{TH},\ldots\}$$

The probabilities of events in $\mathcal{F}_2$ which are also in $\mathcal{F}_1$ have been defined already. For the new events, we use the probability rules: $P(\pi_{HT})=0.5\times0.5=0.25$, $P(\pi_{HH}\cup\pi_{TH})=0.5\times0.5+0.5\times0.5=0.5$, etcetera. Eventually, $(\Omega,\mathcal{F}_2,P/\mathcal{F}_2)$ is fully specified.

And this process of building our filtration gradually goes on. At each step we construct a σ-algebra more refined (larger, inclusion-wise) than the previous ones, and we "reveal" probabilities associated to the freshly-added events.

Before continuing (that is, defining time-dependent random variables $\{X_t\}_{t\in T}$, hence defining the actual process), we make a digression for understanding filtrations in more general terms.

Filtrations as information carriers

We start with an analogy. Suppose Ω (the set of some experiment outcomes) is laid down as a (geographic) map, and x (an event) is a point on this map, of location unknown to us. Let's imagine that a filter $\mathcal{F}_t$ can tell us in which of its sets $U \in \mathcal{F}_t$ resides x. Then $\mathcal{F}_t$ conveys some information about the location of x, although it does not pinpoint x with accuracy. One can visualize $\mathcal{F}_t$, e.g., as a grid on the map, which tells you which patch of the grid contains x, without actually showing x. Then is clear that another filter $\mathcal{F}_T$, with $\mathcal{F}_T \supset \mathcal{F}_t$, will provide more information/knowledge on the whereabouts of x: it reveals smaller sets $V \in \mathcal{F}_T$ for which $x \in V$. In our analogy, the grid on the map is more dense (refined).

In this analogy, the filter 2^{Ω} (of all subsets of Ω) represents a state of total omniscience (we know precisely where x is located), and the filter $\{\varnothing, \Omega\}$ represents a state of total ignorance: doesn't really tell anything about x's whereabouts (except that x *is on the map somewhere*).

Another Analogy: Dyadic Partitions of the Unit Interval

Let $I = [0,1)$ be the unit interval, and $\alpha \in I$. Then α has an unique binary expansion as $\alpha = 0.a_0 a_1 ... a_k ...$, with $\alpha = \sum_{i=0}^{\infty} \frac{a_i}{2^i}$ (representation of α in base 2), and a_i binary digits. Consider the following sequence of *dyadic partitions*:

$$\wp_0 = \{I\}$$

$$\wp_1 = \left\{\left[0, \frac{1}{2}\right), \left[\frac{1}{2}, 1\right)\right\}$$

$$\wp_2 = \left\{\left[0, \frac{1}{4}\right), \left[\frac{1}{4}, \frac{2}{4}\right), \left[\frac{2}{4}, \frac{3}{4}\right), \left[\frac{3}{4}, 1\right)\right\}$$

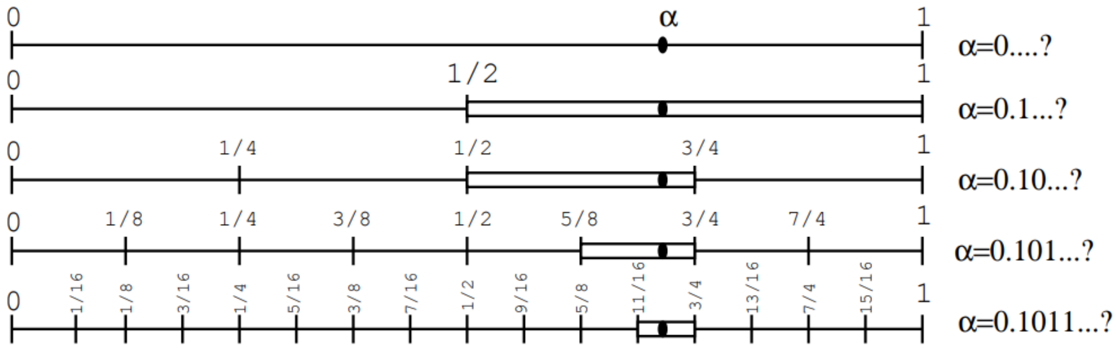
$$\wp_3 = \left\{\left[0, \frac{1}{8}\right), \left[\frac{1}{8}, \frac{2}{8}\right), \left[\frac{2}{8}, \frac{3}{8}\right), \left[\frac{3}{8}, \frac{4}{8}\right), \left[\frac{4}{8}, \frac{5}{8}\right), \left[\frac{5}{8}, \frac{6}{8}\right), \left[\frac{6}{8}, \frac{7}{8}\right), \left[\frac{7}{8}, 1\right)\right\}$$

. . . . .

and suppose α is unknown. Then one can say that

[knowledge of the digits $a_1, a_2, a_3, ..., a_n$ ***]*** $\Leftrightarrow$ ***[knowledge of which element of*** $\wp_n$ ***contains*** α ***]***

This is an illustration:



In other words, $\wp_n$ contains the information about the first n binary digits of α. Note that $\wp_0 \prec \wp_1 \prec \ldots \prec \wp_n \prec \ldots$, in that, every element of $\wp_i$ is included in some element of $\wp_{i-1}$: a finer partition contains more information.

____

Now, back to stochastic processes. Using a tree structure, so far we have defined: (1) the probability space Ω, of all infinite paths in the tree; (2) a filtration: a collection of time-indexed σ-algebras $\{\mathcal{F}_t\}_{t\in T}$ (sometimes called filters), and (3) the probability measure P (defined/revealed gradually, on each filter $\mathcal{F}_t$). The byproducts of this construction are: the "universal filter" $\mathcal{F}_\infty$ (viewed at the smallest σ-algebra containing all $\mathcal{F}_t$'s), and a time-indexed set of probability subspaces $\{(\Omega, \mathcal{F}_t, P/\mathcal{F}_t)\}_{t\in\Gamma}$. We are now prepared to describe the actual process, defined as a time-indexed set of random variables.

Let $X = \{X_t\}_{t\in\Gamma}$ be a generic stochastic process, in the probability space $(\Omega, \mathcal{F}_\infty, P)$. The process X is said to be *adapted to the filtration* $\{\mathcal{F}_t\}_{t\in T}$ if X_t is a well-defined random variable in the probability space $(\Omega, \mathcal{F}_t, P/\mathcal{F}_t)$, for all $t\in\Gamma$ (i.e., X_t is $\mathcal{F}_t$-measurable). We sometimes call X an $\mathcal{F}_\infty$-adapted process, if the connection between $\mathcal{F}_\infty$ and $\{\mathcal{F}_t\}_{t\in T}$ is clearly understood.

Continuing with our example (the binary tree evolution of a system), let's define an adapted stochastic process on the tree:

At time $t_0 = 0$, the system is in an initial/deterministic state x_0^0. This corresponds to the random variable X_0 taking the constant value across Ω: $X_0 : \Omega \to \mathbb{R}$, $X_0(\omega) = x_0$ for all $\omega \in \Omega$ is a well-defined random variable since for $I \subseteq \mathbb{R}$, $X_0^{-1}(I) = \Omega$ if $x_0^0 \in I$, and $X_0^{-1}(I) = \varnothing$ if $x_0^0 \notin I$.

Anticipating a possible martingale property of the process (introduced later), x_0^0 can be viewed as the expected (probability-weighted average) payoff of the process across all of its possible evolutions.

At time t_1, the system can be in one of two possible states, one reached if the first outcome is $w_1^1 = H$ and the other if it is $w_1^2 = T$. These states describe two $\mathcal{F}_1$-measurable sets: $\pi_{w_1^1}$ (or, π_H), and $\pi_{w_1^2}$ (or, π_T). Further, $\pi_{w_1^1}$ represents the set of all evolutions of the process that start with w_1^1 (it <u>filters out</u> all evolutions that start with w_1^2). This justifies the terminology “filter” (or, “filtration”).

We still have to define the random variable X_1 : the process' payoff at time $t=1$. Note that we cannot define X_1 directly (as its domain is infinite). Yet, we make the assumption of knowing the average payoff of the process if it lands in either node, $\widehat{w_1^1}$ or $\widehat{w_1^2}$ (a pedantic notation for $\widehat{H}$ and $\hat{T}$). Assume these values are x_1^1 and x_1^2. This translates into the following averaging expressions:

$$X_1 \text{ average on } \pi_1^1\text{: } x_1^1 = \frac{1}{P\left(\pi_1^1\right)}\int_{\pi_1^1} X_1\left(w\right)dP\left(w\right)\text{, and}$$

$$X_1 \text{ average on } \pi_1^2\text{: } x_1^2 = \frac{1}{P\left(\pi_1^2\right)}\int_{\pi_1^2} X_1\left(w\right)dP\left(w\right).$$

Note that we do not define the function X_1 directly: we rather define it's average on $\mathcal{F}_1$-measurable sets. Obviously, if $x_1^1 \neq x_1^2$, we have $X_1^{-1}\left(x_1^1\right) = \pi_1^1 \in \mathcal{F}_1$, and $X_1^{-1}\left(x_1^2\right) = \pi_1^2 \in \mathcal{F}_1$, and the pre-image of any Borel set is either $\varnothing$, or π_1^1, or π_1^2, or $\pi_1^1 \cup \pi_2^1 = \Omega$. More precisely,

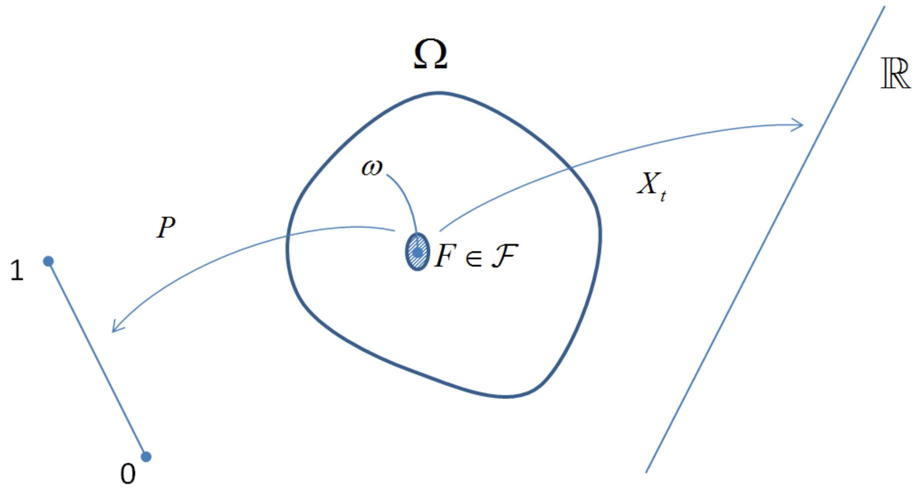
$$X_1^{-1}\left(I\right) = \begin{cases} \pi_1^1 & \text{, if } \quad x_1^1 \in I \quad \text{and } \ x_1^2 \notin I \\ \pi_1^2 & \text{, if } \quad x_1^1 \notin I \quad \text{and } \ x_1^2 \in I \\ \Omega & \text{, if } \quad x_1^1 \in I \quad \text{and } \ x_1^2 \in I \\ \varnothing & \text{, if } \quad x_1^1 \notin I \quad \text{and } \ x_1^2 \notin I \end{cases}$$

All these pre-images are $\mathcal{F}_1$-measurable (which is a pedantic way of saying that they are elements of $\mathcal{F}_1$). Note further that

$$E^P\left[X_1 \mid \mathcal{F}_0\right] = P\left(\pi_1^1\right)\cdot x_1^1 + P\left(\pi_1^2\right)\cdot x_1^2 = \int_{\pi_1^1} X_1\left(w\right)dP\left(w\right) + \int_{\pi_1^2} X_1\left(w\right)dP\left(w\right) = \int_{\Omega} X_1\left(w\right)dP\left(w\right)$$

which is in line with the general definition of an expectation conditioned by a filter. Not that, in general, $E^P\left[X \mid \mathcal{F}_0\right] = E^P\left[X\right]$, as it gives the average of X over Ω.

The main purpose of this illustration is to point out that the random variable X_t is defined indirectly, through its average over a measurable set. This is visualized, generically, in the following figure:



$$\underset{\omega \in F}{\text{average}}\left(X_t(\omega)\right) = \frac{1}{P(F)} \int_F X_t(\omega)\, dP(\omega)$$

Note that the values at the tree's nodes, which in practice stand for some payoffs implied by the process, can be viewed as both the payoff of the process if it reaches a certain node, and as a probability-weighted average, over the infinite set of all possible evolution payoffs of the process that involve that node (evolutions which all start at time $t_0 = 0$). *This is true in general <u>only if the process is a martingale</u>* (notion which will be defined later). Incidentally, this also explains why we are given only two values for the random variable X_1, rather than all the values of X_1 in all the points in Ω. We never define the random variables X_t completely, in the common sense, but rather only their averages across some measurable sets in $\mathcal{F}_t$ - and this is all that we usually need for all practical purposes. Finally, note that two distinct time-indexed random variables of the process are usually defined over different, yet "prefix-related", sets. For example, if X_t and X_{t+1} denote consecutive random variables of our process, then we define X_t as an average over every $\pi_w \in \mathcal{F}_t$, with $|w| = t$ (by $|\cdot|$ we understand the word length), and for each such π_w, we define X_{t+1} over π_{wH} and over π_{wT}, etc. .

Finally, if the process is a martingale (or, driftless - see definition below), one can view any X_t as a time-t-approximation (or, expectation) of a "global" random variable X over the probability space $(\Omega, \mathcal{F}_\infty, P)$: X_t provides the average of X on the measurable sets in $\mathcal{F}_t$. By definition of conditional expectation, we have $\int_A E^P\left[X_t \middle| \mathcal{F}_t\right](\omega)\, dP(\omega) = \int_A X(\omega)\, dP(\omega)$, for all $A \in \mathcal{F}_t$. In this peculiar view, one can very well assume there is only one random variable, X, as under these terms, $X_t = X$ essentially (or, X_t is a *representation* of X at time t). In other words, $X_t(A) = \frac{1}{P(A)} \int_A X(\omega)\, dP(\omega)$ for all $A \in \mathcal{F}_t$. We stress that this view is valid only if the process verifies the following property:

If X is an $\mathcal{F}_\infty$-adapted process, then X is a ***martingale*** if

1. $E^P\left[\,|X_t|\,\right] < \infty$ for all $t \in \Gamma$, and
2. $E^P\left[X_t \mid \mathcal{F}_s\right] = X_s$, for all $t, s \in \Gamma$ with $s < t$.

In particular, $E^P\left[X_t \mid \mathcal{F}_0\right] = X_0$, for all $t \in \Gamma$, which hints at the fact that a martingale is a *driftless* process: the expectation at initial time $t_0 = 0$ of the values of the process at any further time t is never changing (is not drifting), being equal to the initial value X_0.

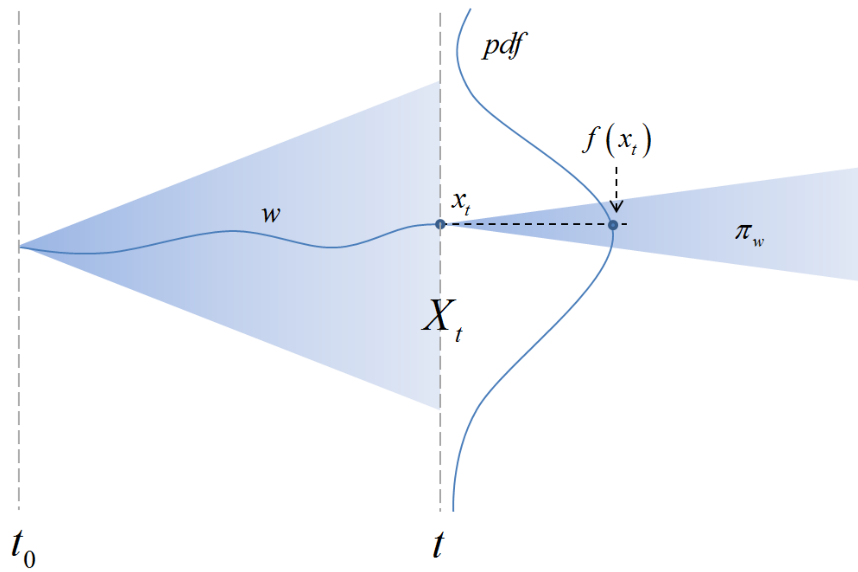
Note carefully that $E^P\left[X_t \mid \mathcal{F}_s\right]$ is a random variable in $\left(\Omega, \mathcal{F}_s, P/\mathcal{F}_s\right)$, which agrees with X_t, in average, on the $\mathcal{F}_s$-measurable sets. Indeed, consider all states of the process X at time s: X_s takes the values $x_s^1, x_s^2, \ldots, x_s^k$, which are averages of X_s over the $\mathcal{F}_s$-measurable sets $\pi_{u_1}, \ldots, \pi_{u_k}$, corresponding to the nodes in the process' tree at depth s. Furthermore, denote by $p_1 = P\left(\pi_{u_1}\right), p_2 = P\left(\pi_{u_2}\right), \ldots, p_k = P\left(\pi_{u_k}\right)$ the probabilities of reaching these nodes, (obviously, $\sum_{i=1}^{k} p_i = 1$, as $\bigcup_{i=1}^{k} \pi_{u_i} = \Omega$). Then the random variable $E^P\left[X_t \mid \mathcal{F}_s\right]$ will take the value $x_s^1 = E^P\left[X_t \mid X_s = x_s^1\right]$ with probability p_1, then the value $x_s^2 = E^P\left[X_t \mid X_s = x_s^2\right]$ with probability p_2, and so on, and the value $x_s^k = E^P\left[X_t \mid X_s = x_s^k\right]$ with probability p_k. Moreover, $E^P\left[X_t \mid \mathcal{F}_t\right] = X_t$, taking the trivial values $x_t^i = E^P\left[X_t \mid X_t = x_t^i\right]$ with associated probabilities.

Further, by the martingale property/definition, one can see that

$$E^P\left[E^P\left[X_t \mid \mathcal{F}_s\right] \mid \mathcal{F}_0\right] = E^P\left[X_s \mid \mathcal{F}_0\right] = X_0 = \sum_{i=1}^{k} p_i \cdot E^P\left[X_t \mid X_s = x_s^i\right] = \sum_{i=1}^{k} p_i \cdot x_s^i .$$

The driftless property of a martingale proves to be very useful in pricing models: if we can describe the evolution of a price (in units of numéraire) of a derivative instrument by a martingale, and if we know all possible terminal payoff values at instrument's maturity, we can compute, working our way backwards, all expectations at each of tree's nodes, reaching the root; that is, we find today's price of the instrument. Prices don't evolve as martingales in the real world; hence most pricing models aim at constructing an equivalent martingale process, based on which they compute today's price values given the payoff at maturity. Basically, one starts with a real-world probability measure, and invoking the fundamental theorem of asset pricing (indirectly enforcing an arbitrage-free environment), it finds an equivalent "risk-neutral" probability under which the price of the asset, in units of numéraire, behaves as a martingale. Once everything is written under the risk-neutral probability measure, the process of backward-evaluating expectations kicks in, allowing to say that the discounted payoff expectation at maturity is the expectation of today's price. This is one important application of change of measure, from a real-world (risky) probability measure to a risk-neutral probability measure. The relationship between the real-world measure and risk-neutral measure is given by the *Fundamental Theorem of Asset Pricing*, and is subject of a later discussion (Section 4.1).

As epilogue to this section we illustrate how one can view a process in continuous time and "space" as an extension of the discrete process that we have described by a binary tree. Consider the following intuitive figure:



We start with a tree-evolution of a process, already described before. Here $t_0 = 0$ is the initial time, t is some specific time in the evolution of the process, X_t is the random variable describing the process state-payoff at time t, w is the history of the process on a particular evolution up to time t, and π_w is the collection of all infinite paths that start with w, that is, a measurable set in $\mathcal{F}_t$. Saying that $X_t(\pi_w) = x_t$ is like saying that the payoff of the process at the end of the path w is x_t, and this is given by $X_t(\pi_w) = \dfrac{1}{P(\pi_w)} \int_{\pi_w} X_t(\omega)\, dP(\omega)$ (in discrete form is simply a weighted average).

Now, if X_t takes values in continuous space (as opposed to discrete), that is, in $\mathbb{R}$, we can view the tree up to level t as a "solid cone" (considering continuous time as well), and we have to replace the probability $P(\pi_w)$ with the *pdf* (probability density function) f of X_t. If we imagine the vertical line at time t to be the real axis ($\mathbb{R}$) and the point where w "touches" the real line as being denoted by x_t, the probability $P(\pi_w)$ of X_t taking the value x_t is replaced by $f(x_t)$, representing informally the likelihood that the payoff at time t will be in the neighborhood of x_t. Then, integrating ("summing up") all $x_t \cdot f(x_t)$ will give us the mean of X_t: $\int_{\mathbb{R}} x_t f(x_t)\, dx_t$. Indeed, intuitively, $x_t f(x_t)$ can be viewed as $\int_{\pi_w} X_t(\omega)\,dP(\omega) = \underbrace{\left(\frac{1}{P(\pi_w)}\int_{\pi_w} X_t(\omega)\,dP(\omega)\right)}_{x_t} \cdot \underbrace{P(\pi_w)}_{dF(x_t)=f(x_t)dx_t} \to x_t f(x_t)\,dx_t$, and integrating these quantities across all paths of length $|w|$, would give us $\int_{\Omega} X_t(\omega)\,dP(\omega) = \int_{\mathbb{R}} x_t f(x_t)\,dx_t$ - which is the known formula for the first moment (expectation) of X_t. This is an intuitive rather than a formal derivation.

2.1.3 Brownian Motion

"If numbers aren't beautiful, I don't know what is."
Paul Erdös, 1913 – 1996

We start with some notation and basic definitions. A *coin flip* is a discrete random variable c taking values in $\{-1,+1\}$, with $c \in \mathcal{U}(0,1)$ (- discrete distribution); that is, is uniformly distributed with mean $E[c]=0$ and variance $E\left[c^2\right]=1$. A *random walk* corresponding to N independent coin flips $\{c_i\}_{i=1}^N$ is the random variable $S_N = \sum_{i=1}^N c_i$, with N being a fixed positive integer. Then, S_N has mean $E[S_N]=0$ and variance $E\left[S_N^2\right]=N$, and approaches a normal distribution when N is getting large.

This is a direct consequence of the *Central Limit Theorem*:

> If $\{c_i\}_{i=1}^N$ is a sequence of independent and identically distributed random variables, with $E[c_i]=\mu$ and $Var[c_i]=\sigma^2$, then as N approaches ∞, the random variable $\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N c_i - \mu\right)$ converges in distribution to $\mathcal{N}(0,\sigma^2)$. Consequently, the distribution of the sample mean $\frac{\sum_{i=1}^N c_i}{N}$ is approximated by $\mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right)$ when N is becoming large.

If we replace the sample mean with the random walk S_N already defined, we obtain that $\frac{S_N}{\sqrt{N}}$ converges in distribution to $\mathcal{N}(0,1)$ when $N \to \infty$. We write $\frac{S_N}{\sqrt{N}} \xrightarrow[N\to\infty]{d} \mathcal{N}(0,1)$. Of our main interest is the limit process $\lim_{N\to\infty}\frac{S_N}{\sqrt{N}}$, as we'll see that $\frac{S_N}{\sqrt{N}}$ is the discrete equivalent of a standard Brownian motion.

We now make a transition from discrete time steps to continuum. Let $[0,T]$ be a time interval and $\Delta:\quad 0=t_0<t_1<\ldots<t_N=T$ be an equally spaced division, with $\Delta t=t_{i+1}-t_i=\frac{T}{N}$. When needed, we will perform the following conversions:

from continuum to discrete: $t\to i_t=\left\lfloor t\frac{N}{T}\right\rfloor$*, and from discrete to continuum:* $i\to t_i=i\frac{T}{N}$.

Now denote $x_t=\sqrt{\frac{T}{N}}\cdot S_{\left\lfloor t\frac{N}{T}\right\rfloor}$, where $S_{\left\lfloor t\frac{N}{T}\right\rfloor}=S_{i_t}$ is a random walk of i_t steps. The process x_t is the <u>discrete</u> version of a *standard Brownian motion*. We immediately have that $E[x_t]=E\left[S_{i_t}\right]=0$. Moreover,

$$x_t=\sqrt{\frac{T}{N}}\cdot S_{\left\lfloor t\frac{N}{T}\right\rfloor}=\sqrt{\frac{T}{N}}\cdot\frac{S_{\left\lfloor t\frac{N}{T}\right\rfloor}}{\sqrt{\left\lfloor t\frac{N}{T}\right\rfloor}}\cdot\sqrt{\left\lfloor t\frac{N}{T}\right\rfloor}=\frac{S_{i_t}}{\sqrt{i_t}}\cdot\sqrt{t}\xrightarrow[i_t\to\infty,\text{ or }N\to\infty]{d}\mathcal{N}(0,1)\cdot\sqrt{t}=\mathcal{N}(0,t)\ ,$$

in other words, as N approaches ∞, x_t is normally distributed with mean 0 and variance t (that is, $E[x_t]=0$, $E\left[x_t^2\right]=t$). A *standard Brownian motion* is the limit process x_t, when $N\to\infty$.

Since the random walk S_{i_t} is a Markov process (memory-less), it follows that x_t is a Markov process as well: $\{x_t-x_s\}_{s,t}$ are mutually independent and have the same distribution. Then, x_t-x_s has the same distribution as x_{t-s}, meaning that $x_{t+\Delta t}-x_t=\Delta x_t\in\mathcal{N}(0,\Delta t)$: x_t has a Gaussian step too.

More formally, and defined by properties rather than construction, a *standard Brownian* motion (*standard BM* for short) is a continuous-time stochastic process $\{x_t\}_{t\in\mathcal{T}}$ with the following properties:

1. $x_0=0$
2. the mapping $t\to x_t$ is almost surely (with probability 1) continuous:

   $$\lim_{t\to t_0}E\left[\left(x_t-x_{t_0}\right)^2\right]=0,\ \text{for all } \mathrm{t}_0\in\Gamma$$ (Γ is the usual time-index set)
3. $\{x_t\}_{t\in\Gamma}$ has stationary, independent increments:

   (a) $x_{s+t}-x_s$ has the same distribution as $x_t=x_t-x_0$

(b) $\left\{x_{t_i} - x_{t_j}\right\}$ are jointly independent

4. $x_{s+t} - x_s$ is in $\mathcal{N}\left(0,t\right)$: the process exhibits a Gaussian step

We have created a standard BM in the limit, $\left\{x_t\right\}_{t\in\Gamma}$, from a coin flip and a random walk, with $x_t \in \mathcal{N}\left(0,t\right)$ and $\Delta x_t \in \mathcal{N}\left(0,\Delta t\right)$. In other words, $x_t \in \mathcal{N}\left(0,1\right)\sqrt{t}$, $\Delta x_t \in \mathcal{N}\left(0,1\right)\sqrt{\Delta t}$, allowing us to simulate a standard BM by drawing a value from a standard normal distribution, $\phi_t \in \mathcal{N}\left(0,1\right)$ and setting $x_t = \phi_t\sqrt{t}$, or $\Delta x_t = \phi_t\sqrt{\Delta t}$, whichever is required.

Note. A standard Brownian motion is a

Gaussian process,

diffusion process,

Markov process,

martingale,

fractal (statistically self-similar).

Intuitively, a statistical self-similar (fractal) process is invariant in distribution under a suitable scaling of time and space. More rigorously: a stochastic process $\left\{x_t\right\}_{t\in\Gamma}$ is self-similar if for any $a>0$ there exists $b>0$ such that x_{at} and bx_t are equal in distribution. It can be shown that when this happens, there exists an unique $H \geq 0$ such that $b = a^H$ (H is called the Hurst exponent).

A standard BM is indeed self-similar: it can be seen that $\left\{y_t = x_{at}\right\}_{t\in\mathcal{T}}$ is also a standard BM with mean zero. Moreover, $Var\left[a^{\frac{-1}{2}}y_t\right] = t$, hence setting $H = \frac{-1}{2}$, the process $\left\{z_t = a^H x_t\right\}_{t\in\mathcal{T}}$ has the same distribution as y_t.

To convey more intuition, consider $a = \frac{1}{3}$, and that the time universe (where t resides) along which x evolves is <u>compressed</u> into another universe (represented by $t' = \frac{1}{3}t$) so that whatever x "accomplishes" in a certain length of time t, y accomplishes a-times faster: in $t' = \frac{1}{3}t$. Then, compressing the time by a factor $\frac{1}{3}$ results in <u>expanding</u> the "space" by a factor of $\sqrt{\frac{1}{3}}$:

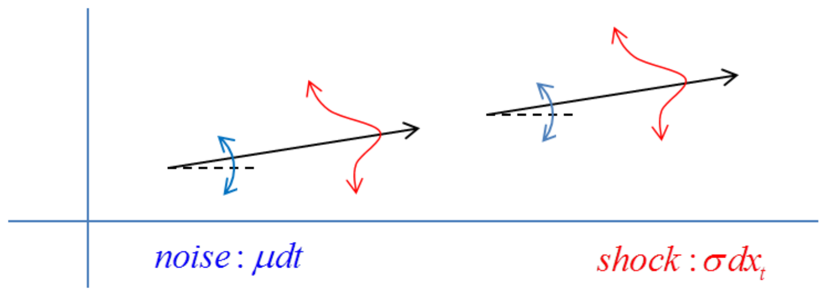
$z_t = \frac{x_t}{\sqrt{1/3}} = \sqrt{3}x_t$. By abuse of notation, the processes $x_{\frac{1}{3}t}$ and $\sqrt{3}x_t$ have the same distribution, meaning that they are statistically equivalent.

The standard BM $\{x_t\}_t$ has a zero <u>*drift*</u> *(rate of change in expectation)* and <u>*diffusion*</u> *1 (square root of rate of change in variance)*: $\left(\mathrm{E}\left[x_t\right]\right)'_t = (0)'_t = 0$ and $\sqrt{\left(\mathrm{E}\left[x_t^{\,2}\right]\right)'_t} = \sqrt{(t)'_t} = 1$ (here the derivatives are obviously taken with respect to time).

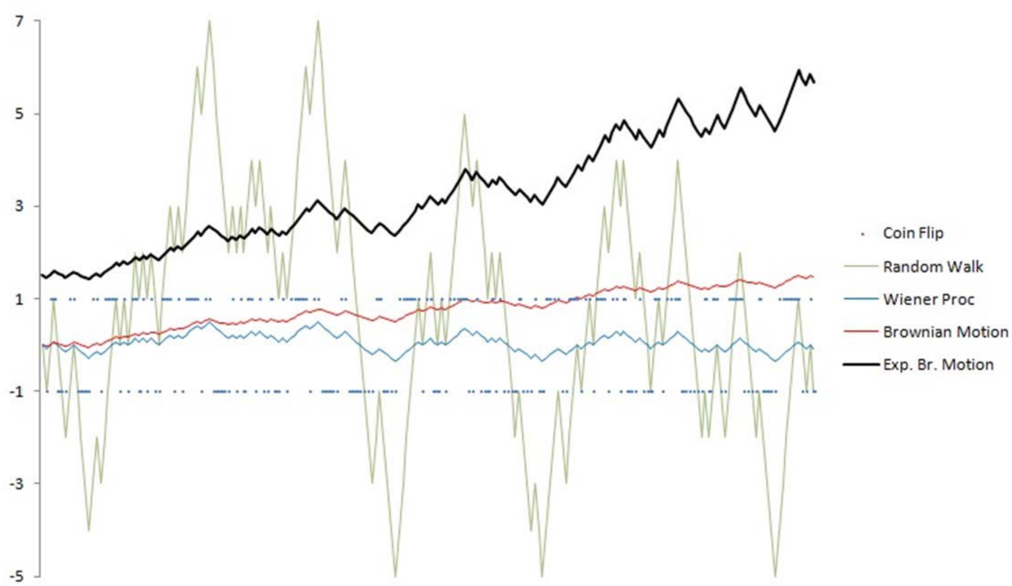
To make the model more realistic, we want to consider a process with an arbitrary drift μ and an arbitrary diffusion σ^2: define the process $\{X_t\}_t$ by the equation $X_t = \mu t + \sigma x_t$, or equivalently, $dX_t = \mu dt + \sigma dx_t$. We have: $E\left[X_t\right] = \mu t$ and $Var\left[X_t\right] = \sigma^2 t$, hence $X_t \in \mathcal{N}\left(\mu t, \sigma^2 t\right)$. $\{X_t\}_t$ is called a *generalized Brownian motion*, or simply *Brownian motion* (BM for short) with drift μ and diffusion σ. The mean and variance of this process increase linearly in time. This is a simulation of such process:

The Brownian motion is suitable for modeling stochastic behavior of equity prices, however, in terms of returns and not of prices themselves. We therefore use it to simulate a stock's *instantaneous rate of return,* as follows. If P_t is the price of a stock at time t, then $\frac{dP_t}{P_t}$ is stock's instantaneous rate of return, given by the Brownian motion $\frac{dP_t}{P_t} = \mu dt + \sigma dx_t$, that is, we simulate returns rather than prices as a BM, where x_t is a *standard BM*. Rewriting, $dP_t = \mu P_t dt + \sigma P_t dx_t$. We will show later (as application of Itô's Lemma) that the solution to this PDE is $P_t = P_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma x_t}$, with P_0 being the initial price; in other words, the stock price is a random variable distributed lognormally: $\ln\left(P_t\right)=\ln\left(P_0\right)+\left(\mu-\frac{\sigma^2}{2}\right)t+\sigma x_t$ with $\ln\left(P_t\right)$ normally distributed.

A process that follows the stochastic differential equation (SDE for short) $dP_t=\mu P_t dt+\sigma P_t dx_t$, with x_t a BM, is called a *Geometric Brownian Motion*, GBM for short. A GBM is a combination of noise and shock:



We close this discussion, illustrating the processes that we went through: from tossing a coin, to simulating a random walk, then standard Brownian motion, a generalized Brownian motion, and finally a geometric Brownian motion:



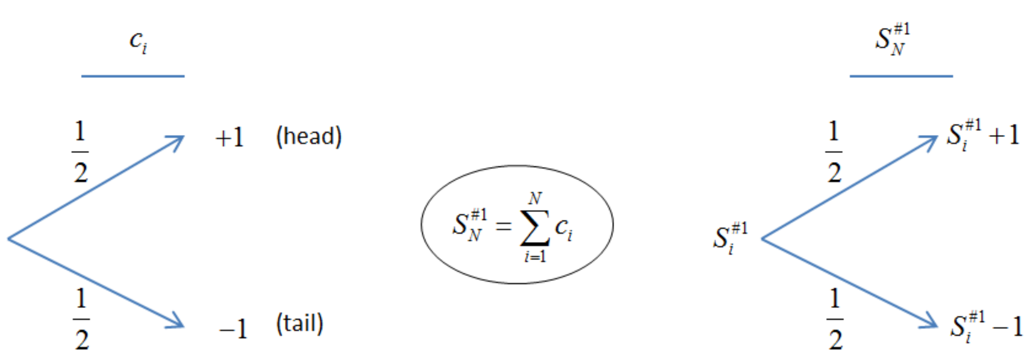
Standard BM is also known as a *Wiener process*, and the GBM as *exponential Brownian motion* (see the legend above). Note that only one source of randomness (a coin toss) was used to produce all the above plots: except for the coin toss (plotted in small grey dots), all the others have been derived deterministically, one from the previous one, in the order given by the legend. Note that the GBM starts looking like a stock price chart, and indeed looks very much like one when the time division approaches zero.

We end this section with a note on filtrations induced by Brownian motions (or any other stochastic process, for that matter). Let $\{X_t\}_{t\geq 0}$ be a Brownian motion process defined on some probability space $(\Omega, \mathcal{F}, P)$. For any given t, X_t is a "standalone" random variable, $X_t : \Omega \to \mathbb{R}$, and recall (Section 2.1.1) that the collection of pre-images of all Borel sets $\{X_t^{-1}(B) | B \text{ is a Borel set}\}$ is a sub-σ-algebra of $\mathcal{F}$: we call it the σ-algebra induced by X_t. Consider now the following σ-closure: $\mathcal{F}_t = \sigma\left[\bigcup_{0\leq s\leq t}\{X_s^{-1}(B) | B \text{ is a Borel set}\}\right]$, for all $t \geq 0$. Is easy to see that $\{\mathcal{F}_t\}_t$ is a legitimate filtration over Ω. $\{\mathcal{F}_t\}_t$ is said to be induced by the process $\{X_t\}_t$. Obviously, $\{X_t\}_t$ is adapted to the filtration $\{\mathcal{F}_t\}_t$, as clearly X_t is $\mathcal{F}_t$-measurable for all $t \geq 0$.

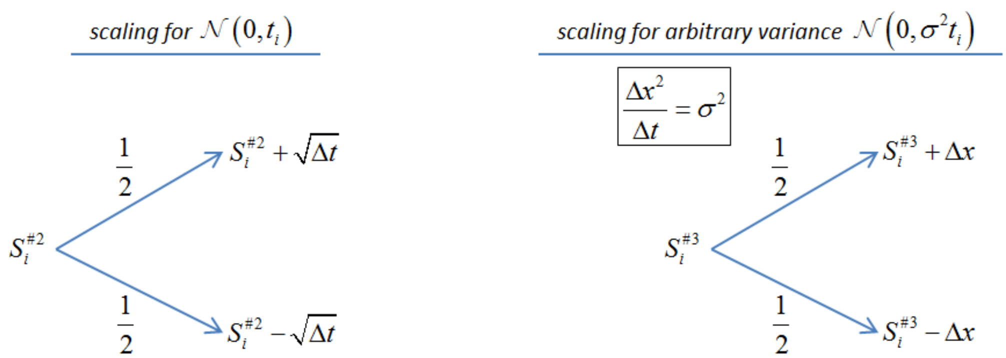
Biased Random Walks

This section is particularly important as it connects with the derivation of the Fokker-Planck equation, as laid out shortly after. In the following, we will use a superscript, such as in $S^{\#k}$, to distinguish among different processes that will be discuss. The superscripts will reflect the order in which the processes are introduced to the reader.

Previously, we constructed Brownian motions, in the limit, from a discrete random walk governed by an unbiased coin toss (with an equal probability of occurrence of heads and tails). Let's recall the construction of a discrete Brownian motion. We have started with an equally spaced division $\Delta: \quad 0 = t_0 < t_1 < ... < t_N = T$, of a time interval $[0, T]$, and an unbiased coin. Denote $\Delta t = \frac{T}{N}$, and set the initial state of a process $\{S_i^{\#1}\}_{i\in\{0,...,N\}}$ as $S_0^{\#1} = 0$ (the point of origin for our random walk). At each time step $t_i = i\Delta t$, we toss the unbiased coin and advance the state of the process to either $S_{i+1}^{\#1} = S_i^{\#1} + 1$ or $S_{i+1}^{\#1} = S_i^{\#1} - 1$, depending on the outcome of the toss - heads or tails, respectively (the outcome of the i^{th} toss is denoted by c_i):



Note that, more rigorously, in the above figure (left side), we rather depict the payoff of the coin toss, that is, a random variable. Obviously, $S_i^{\#1} = \sum_1^i c_i$, with mean zero and variance i. When N approaches infinity, the time step approaches zero, the random variable $\frac{S_N^{\#1}}{\sqrt{N}}$ approaches in distribution $\mathcal{N}(0,1)$, and the standard Brownian motion W_{t_i} is approximated by $\sqrt{t_i}\frac{S_i^{\#1}}{\sqrt{i}}$, distributed as $\mathcal{N}(0,t_i)$. Here, the connection between times and steps is $t_i = i\Delta t = i\frac{T}{N}$. Note the scaling that we applied to $S_N^{\#1}$: the division by $\sqrt{N}$, or $\sqrt{i}$ respectively, in order to enforce the unit variance. We can remove this scaling by changing the magnitude of the random walk step as follows:



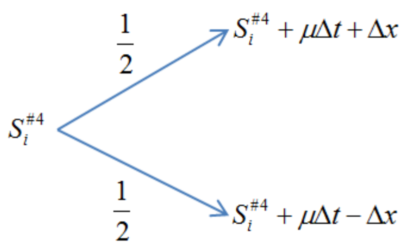
In the figure to the left, the newly transformed process $S_i^{\#2}$ is given by the expression

$$S_i^{\#2} = \sqrt{\Delta t}\sum_1^i c_i = \sqrt{i\cdot\Delta t}\frac{S_i^{\#1}}{\sqrt{i}} = \sqrt{t_i}\frac{S_i^{\#1}}{\sqrt{i}} \xrightarrow{\Delta t\to 0} W_{t_i} \in \mathcal{N}(0,t_i).$$

And we can do even better: we can induce an arbitrary diffusion σ , as in the figure to the right. Indeed, in the right picture, we have $\Delta x = \sigma\sqrt{\Delta t}$ and $S_i^{\#3} = \sigma\sqrt{\Delta t}\sum_1^i c_i \xrightarrow{\Delta t\to 0} \overline{W_{t_i}} \in \mathcal{N}(0,\sigma^2 t_i)$, where $\overline{W_{t_i}}$ is a Brownian motion standing for σW_{t_i} . Note that there is an intimate relationship between the step length and the step duration: $(\Delta x)^2$ is of the same order as Δt : $\frac{(\Delta x)^2}{\Delta t} = \sigma^2$, that is, a constant. This property will become more apparent in Section 2.2.

So far we have learned how to tweak the random walk step in order to change the diffusion of the limiting process. How about the drift? Adding some drift to our new random walk is even a simpler matter:

unbiased random walk with drift

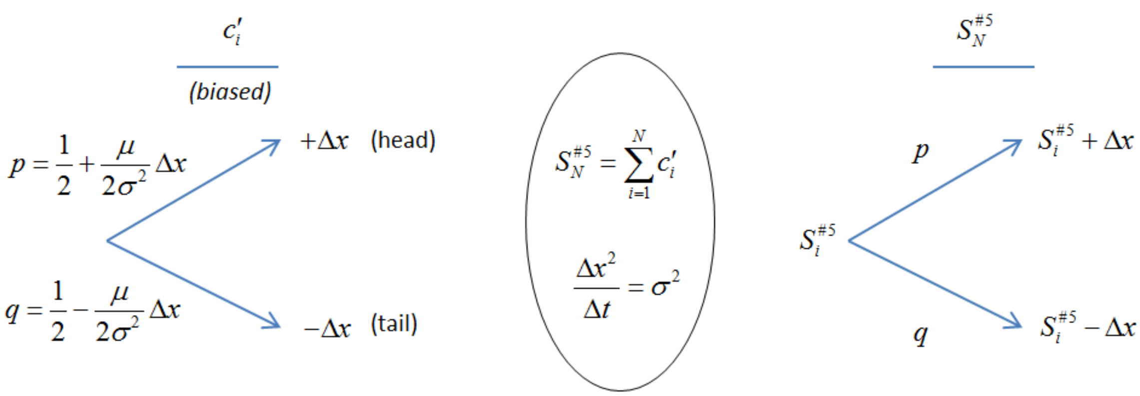


Indeed, the above process can be written as:

$$S_i^{\#4} = \sum_1^i \left(\mu\Delta t + c_j\Delta x\right) = \mu\left(i\Delta t\right) + \sigma\sqrt{i\Delta t}\,\frac{\sum_1^i c_i}{\sqrt{i}},$$

hence in the limit, $S_i^{\#4}$ converges to $X\left(t_i\right) = \mu t_i + \sigma W\left(t_i\right)$, where $W \in \mathcal{N}\left(0, t_i\right)$ is a standard Brownian motion. Thus, the process $S^{\#4}$ belongs (in the limit) to $\mathcal{N}\left(\mu t_i, \sigma^2 t_i\right)$.

So far, all discussed random walks have been unbiased, that is, governed by an unbiased coin. Our aim is to construct a <u>biased</u> random walk that exhibits the same distribution (in particular, same first two moments) as $S^{\#4}$, with symmetric payoff, and which converges in the limit to X. We claim that the following random walk does the trick:



The processes $S^{\#4}$ and $S^{\#5}$ are equivalent (for our purpose), in that they both approximate in the limit the process $X\left(t_i\right) = \mu t_i + \sigma W\left(t_i\right)$, where W is a standard Brownian motion. The discrete process $S^{\#5}$ will be used later to "represent" the continuous process given by the SDE $dX_t = \mu dt + \sigma dW_t$ and $X_0 = 0$. Let's now justify these claims.

From now on, let's agree to drop the superscript in $S^{\#5}$, and the apostrophe in c' (while remembering that the coin is biased this time, governed by probabilities p and q). First, as any random walk, we know that S must exhibit a normal distribution of some sort, in the limit, due to the Central Limit Theorem (see Section 2.1.3). We also know that $S_i = \sum_{j=1}^{i} c_j$, and that $\sqrt{i}\left(\frac{1}{i}\sum_{j=1}^{i} c_j - E[c]\right)$ converges in distribution to $\mathcal{N}(0, Var[c])$. Here we denoted by c the payoff of one generic coin (they are all iid). The two moments of c are given by

$$E[c] = p\Delta x - q\Delta x = \mu\frac{\Delta^2 x}{\sigma^2} = \mu\Delta t \text{ , and}$$

$$Var[c] = E\left[c^2\right] - E[c]^2 = \Delta^2 x - (\mu\Delta t)^2 \text{ .}$$

By Central Limit Theorem, we have that $S_i = \sum_{j=1}^{i} c_j$ converges in distribution to $\mathcal{N}(iE[c], iVar[c])$.

Now, $iE[c] = i\mu\Delta t = \mu t_i$ and $iVar[c] = i\Delta^2 x - i(\mu\Delta t)^2 = i\sigma^2\Delta t - \mu^2\left(i\Delta^2 t\right) \overset{t_i = i\Delta t}{=} \sigma^2 t_i - t_i\Delta t\mu^2$. Then, observe that $t_i\Delta t\mu^2 \to 0$ when $\Delta t \to 0$. Thus S_i converges in distribution to $\mathcal{N}\left(\mu t_i, \sigma^2 t_i\right)$.

For a better understanding of this process, let's provide an alternative derivation of the first two moments of S_i. This will rather be a combinatorial derivation: we will not use Central Limit Theorem. Let's concentrate our attention on this last process alone:

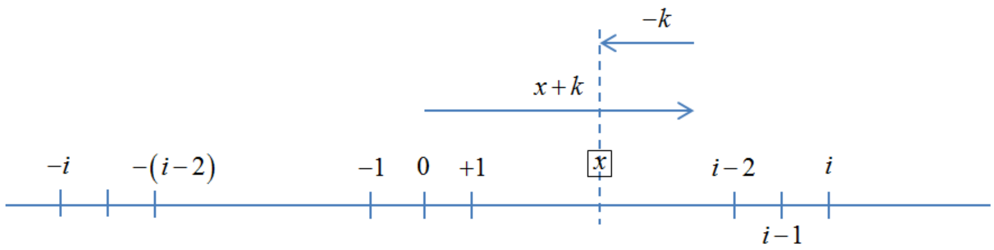
biased random walk

$$p = \frac{1}{2} + \frac{\mu}{2\sigma^2}\Delta x \quad S_i \to S_i + \Delta x$$

$$q = \frac{1}{2} - \frac{\mu}{2\sigma^2}\Delta x \quad S_i \to S_i - \Delta x$$

where: $\begin{cases} \Delta t = \dfrac{T}{N} \\ i = \dfrac{t_i}{\Delta t} \\ \dfrac{\Delta x^2}{\Delta t} = \sigma^2 \end{cases}$

First, let's consider a simpler random walk $\overline{S_i}$, where the advancement in either direction is one unit of length (rather than Δx), and ask what is the probability of reaching a certain distance x from the origin (zero), after i steps:



Obviously, the farthest reachable points are $+i$ and $-i$, anything farther is unreachable (thus reached with probability zero). Also, not all closer points can be reached; for example, $i-1$ cannot be reached. Yet, $i-2$ can be reached, say, by stepping forward $i-1$ times and step backwards once. In general, a point x can be reached only if $2 \mid i-x$ (i.e., $i-x$ is an even number); in other words, $i-x=2k$ for some integer k: the walk consists of $x+k$ steps in the direction of x and k steps in the opposite direction (as shown in the above figure). This means that $\mathrm{Prob}\left(\overline{S_i}=x\right)>0$ only for $x \in \left\{-i,-(i-2),\ldots,(i-2),i\right\}$. For these points, $\mathrm{Prob}\left(\overline{S_i}=x\right)$ is the probability of i Bernoulli trials resulting in exactly $x+k=i-k$ successes, given by $\mathrm{Prob}\left(\overline{S_i}=x\right)=C_i^{i-k}p^{i-k}q^k$. For brevity, we denote this probability as $v_{x,i}$ and we express it in terms of i and x:

$$v_{x,i}=C_i^{\frac{1}{2}(i+x)}\cdot p^{\frac{1}{2}(i+x)}q^{\frac{1}{2}(i-x)} \ ,$$

which is essentially a binomial distribution with the first moments given by

$$E\left[\overline{S_i}\right]=\sum_{x=-i}^{+i}x\cdot v_{x,i}=\sum_{k=0}^{i}(2k-i)v_{2k-i,i}=\sum_{k=0}^{i}(2k-i)\cdot C_i^k p^k q^{i-k}=i(p-q) \ ,$$

$$Var\left[\overline{S_i}\right]=\sum_{k=0}^{i}(2k-i)^2 v_{2k-i,i}=\sum_{k=0}^{i}(2k-i)^2 C_i^k p^k q^{i-k}=4ipq \ .$$

To reconstruct S_i from $\overline{S_i}$ we change back the step: from unit step to a fractional step Δx, which has the effect of scaling the expectation by Δx and variance by Δx^2 (since $S_i=\Delta x\overline{S_i}$), and we replace $i=\frac{t_i}{\Delta t}$ and $\frac{(\Delta x)^2}{\Delta t}=\sigma^2$, thus obtaining the moments of our original random walk S_i:

$$E\left[S_i\right]=\frac{t_i}{\Delta t}(p-q)\Delta x \overset{p-q=\frac{\mu}{\sigma^2}\Delta x}{=} \frac{t_i}{\Delta t}\cdot\frac{\mu}{\sigma^2}\cdot(\Delta x)^2 \overset{\Delta x^2=\sigma^2\Delta t}{=} \mu t_i \text{ , and}$$

$$Var\left[S_i\right] = 4pqt_i \frac{(\Delta x)^2}{\Delta t} \overset{\frac{(\Delta x)^2}{\Delta t}=\sigma^2}{=} 4t_i \left[\left(\frac{1}{2}\right)^2 - \underbrace{\left(\frac{\mu}{2\sigma^2}\Delta x\right)^2}_{=\frac{\mu^2 \Delta t}{2} \to 0} \right] \sigma^2 \overset{\Delta t \to 0}{\to} t_i \sigma^2 \ ,$$

as expected. Then, when $\Delta t \to 0$ (or equivalently, $N \to \infty$), the binomial distribution approaches the normal distribution with expected value μt_i and variance $\sigma^2 t_i$. Translating the notation $v_{x,i} = \text{Prob}\left(\overline{S_i} = x\right)$ into S_i-terms: $v_{y,j} = \text{Prob}\left(\underbrace{S_{j\Delta t}}_{t_j} = \underbrace{y\Delta x}_{z} \right) = \text{Prob}\left(S_{t_j} = z\right)$, we then have

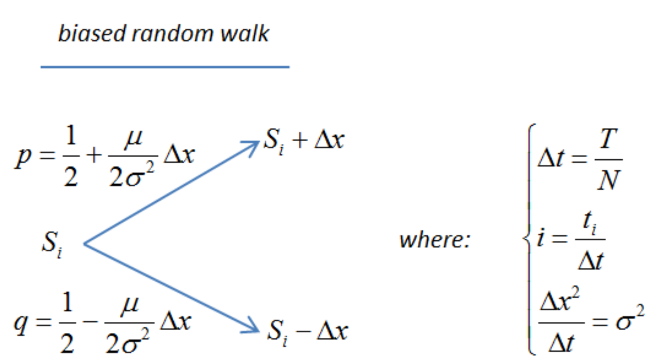
$$v_{y,j} \overset{N\to\infty}{\to} f\left(z, t_j\right) = \frac{1}{\sqrt{2\pi\sigma^2 t_j}} e^{-\frac{\left(z-\mu t_j\right)^2}{2\sigma^2 t_j}} \ .$$

Note that the values $v_{y,j} = f\left(y\Delta x, j\Delta t\right)$ cover more and more points in the plane as Δx and Δt approach zero. In the limit, the density function f is fully defined.

Fokker-Plank Equation

Let's consider the class of processes which rely on a standard Brownian motion (W_t) as a source of randomness (entropy), namely, on processes described by the SDE $dX_t = \mu dt + \sigma dW_t$, with $X_0 = 0$; or equivalently, $X_t = \mu t + \sigma W_t$. The drift of X_t is μ (i.e., $E\left[X_t\right] = \mu t$) and the diffusion is σ (i.e., $Var\left[X_t\right] = \sigma^2 t$). We have already made the case that X_t is the limit of the discrete process S_i, previously constructed. Using S_i as a proxy for X_t, here we answer the following question: how does the density function of such process change in time?

Let's refresh the anatomy of S_i :



We have a time horizon T, an equally spaced division $\Delta: \quad 0 = t_0 < t_1 < \ldots < t_N = T$ of the time interval $\left[0, T\right]$, and define $\Delta t = \dfrac{T}{N}$, $t_i = i\Delta t$. The process S_i advances one step at a time, by tossing a biased coin (governed by probabilities p and q) and changing its state by $\pm\Delta x$ based on the toss outcome: biased heads or biased tails. The initial state of the process $\left\{S_i\right\}_{i\in\{0,\ldots,N\}}$ is $S_0 = 0$. When $\Delta t \to 0$ (or equivalently, $N \to \infty$), the binomial distribution approaches the normal distribution with expected value μt_i and variance $\sigma^2 t_i$:

$$v_{x,i} \overset{\Delta t \to 0}{\to} f\left(x, t_i\right) = \frac{1}{\sqrt{2\pi\sigma^2 t_i}} e^{-\frac{\left(x-\mu t_i\right)^2}{2\sigma^2 t_i}} .$$

where we reused the notation $v_{x,t} = \mathrm{Prob}\left(S_i = x\right)$. We seek to derive a PDE that describes f. We start with the discrete case and observe that we can reach the state $S_{i+1} = x$ from the state $S_i = x - \Delta x$ with probability p, and from the state $S_i = x + \Delta x$ with probability q, which leads to the difference equation:

$$v_{x,i+1} = p v_{x-\Delta x,i} + q v_{x+\Delta x,i} .$$

In the continuous case, this translates into

$$f\left(x, t+\Delta t\right) = p\, f\left(x-\Delta x, t\right) + q\, f\left(x+\Delta x, t\right) .$$

Expressing each term as a Taylor expansion around the point $\left(x, t+\Delta t\right)$, and considering that $\left(\Delta x\right)^3$, $\Delta t \Delta x$,... approach zero faster than Δt, and keeping only the Δt and Δx^2 terms, we obtain:

$$f\left(x, t+\Delta t\right) = f\left(x,t\right) + \Delta t \frac{\partial f}{\partial t}\left(x,t\right) + \ldots$$

$$f\left(x+\Delta x, t\right) = f\left(x,t\right) + \Delta x \frac{\partial f}{\partial x}\left(x,t\right) + \frac{\left(\Delta x\right)^2}{2}\frac{\partial^2 f}{\partial x^2}\left(x,t\right) + \ldots$$

$$f\left(x-\Delta x, t\right) = f\left(x,t\right) - \Delta x \frac{\partial f}{\partial x}\left(x,t\right) + \frac{\left(\Delta x\right)^2}{2}\frac{\partial^2 f}{\partial x^2}\left(x,t\right) + \ldots$$

and substituting in the initial expression for f, we obtain

$$\Delta t \frac{\partial f}{\partial t}\left(x,t\right) = \left(q-p\right)\Delta x \frac{\partial f}{\partial x}\left(x,t\right) + \frac{\left(\Delta x\right)^2}{2}\frac{\partial^2 f}{\partial x^2}\left(x,t\right)$$, and observing that $q - p = -\dfrac{\mu}{\sigma^2}\Delta x$, further dividing by Δt and making the necessary substitutions, we obtain:

$$\boxed{\frac{\partial f}{\partial t}(x,t) = -\mu \cdot \frac{\partial f}{\partial x}(x,t) + \frac{\sigma^2}{2} \cdot \frac{\partial^2 f}{\partial x^2}(x,t)}$$

which is the Fokker-Planck equation describing the evolution of the probability density function for X_t. Let's check that indeed, the equation is correct (is verified by the normal distribution):

normal distribution: $f(x,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-\mu t)^2}{2\sigma^2 t}}$

$$\frac{\partial f}{\partial t}(x,t) = -\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi\sigma^2 t^3}} e^{-\frac{(x-\mu t)^2}{2\sigma^2 t}} + \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-\mu t)^2}{2\sigma^2 t}} \cdot \frac{2\mu(x-\mu t)\cdot 2\sigma^2 t + (x-\mu t)^2 \cdot 2\sigma^2}{\left(2\sigma^2 t\right)^2}$$

$$\frac{\partial f}{\partial x}(x,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-\mu t)^2}{2\sigma^2 t}} \cdot \frac{-2}{2\sigma^2 t}(x-\mu t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-\mu t)^2}{2\sigma^2 t}} \cdot \frac{-(x-\mu t)}{\sigma^2 t}$$

$$\frac{\partial^2 f}{\partial x^2}(x,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \left[e^{-\frac{(x-\mu t)^2}{2\sigma^2 t}} \cdot \left(\frac{-(x-\mu t)}{\sigma^2 t} \right)^2 + e^{-\frac{(x-\mu t)^2}{2\sigma^2 t}} \cdot \frac{-1}{\sigma^2 t} \right]$$

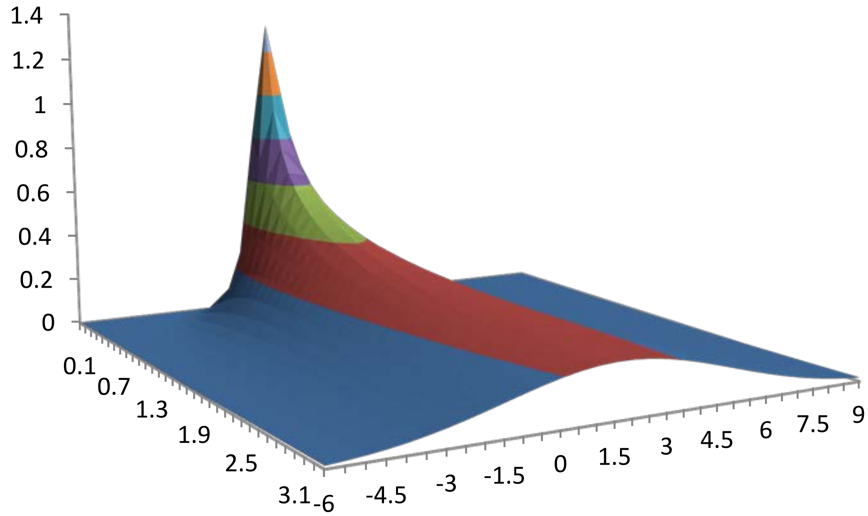
Denoting $\alpha = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-\mu t)^2}{2\sigma^2 t}}$, we obtain:

$$\underbrace{\frac{\partial f}{\partial t}(x,t)}_{LHS} = \alpha \frac{2\mu t(x-\mu t) + (x-\mu t)^2 - \sigma^2 t}{2\sigma^2 t^2}$$

$$\underbrace{-\mu \cdot \frac{\partial f}{\partial x}(x,t) + \frac{\sigma^2}{2} \cdot \frac{\partial^2 f}{\partial x^2}(x,t)}_{RHS} = \alpha \left(\frac{\mu(x-\mu t)}{\sigma^2 t} + \frac{\sigma^2}{2} \cdot \left(\frac{-(x-\mu t)}{\sigma^2 t} \right)^2 + \frac{\sigma^2}{2} \cdot \frac{-1}{\sigma^2 t} \right) =$$

$$= \alpha \left(\frac{\mu(x-\mu t)}{\sigma^2 t} + \frac{(x-\mu t)^2}{2\sigma^2 t^2} - \frac{1}{2t} \right) = LHS \qquad \square$$

The following figure shows a transition density governed by the Fokker-Planck equation just derived. The density evolves in time, from a Dirac delta function, by shifting its mean linearly and decaying its convexity quadratically:


1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.1
0.7
1.3
1.9
2.5
3.1
-6
-4.5
-3
-1.5
0
1.5
3
4.5
6
7.5
9

2.2 Itô's Formula

"Il ne suffit pas de connaître la vérité, il faut encore la proclamer."
Louis Pasteur, 1822 – 1895

Given two processes X and Y, we denote by $\langle X\rangle_T$ the *quadratic variation* of X in the interval $[0,\mathrm{T}]$, and by $\langle X,Y\rangle_T$ the *covaria<u>tion</u>* of X and Y in the same interval (not to be confused with covaria<u>nce</u>, although the connection between these notions will become apparent shortly). More precisely, if $\Delta=\{0=t_0,t_1,...,t_n=T\}$ is an equally spaced division of the time interval $[0,T]$, with norm $\|\Delta\|$, then

$$\langle X,Y\rangle_T=\lim_{\|\Delta\|\to 0}\sum_{\Delta}E\left[\left(X_{t_{i+1}}-X_{t_i}\right)\left(Y_{t_{i+1}}-Y_{t_i}\right)\right]\quad,\quad\langle X\rangle_T=\langle X,X\rangle_T=\lim_{\|\Delta\|\to 0}\sum_{\Delta}E\left[\left(X_{t_{i+1}}-X_{t_i}\right)^2\right]$$

As a note, it is intuitively clear that $d\langle X\rangle_T=E\left[d^2X\,|\,\mathcal{F}_T\right]$ and $d\langle X,Y\rangle_T=E\left[dX\,dY\,|\,\mathcal{F}_T\right]$. To see why, note that $\Delta\langle X\rangle_T$ is a difference of two sums $\sum_{\Delta}...-\sum_{\Delta}...$, that is, $d\langle X\rangle_T\simeq\langle X\rangle_{T+\Delta t}-\langle X\rangle_T$, and only the "last term" of one sum, of the form $\left(X_{t_{i+1}}-X_{t_i}\right)^2$, survives. That term can be viewed as Δ^2X, and at the limit, we get the sought equality.

These concepts apply to deterministic functions of time as well. If x is a well-behaved <u>deterministic</u> function of time, its quadratic variation is zero. Indeed, consider the above-defined division Δ of the time interval $[0,T]$ (any time-interval $[a,b]$ would work), and denote $\Delta t=t_{i+1}-t_i$. Then we can write

$$\langle x\rangle_T=\sum_{i=1}^{n}\left(x(t_{i+1})-x(t_i)\right)^2\overset{MVT}{=}\sum_{i=1}^{n}\left(x'(\xi_i)\Delta t\right)^2=\Delta t\sum_{i=1}^{n}x'(\xi_i)^2\,\Delta t\xrightarrow[\text{or, }n\to\infty]{\Delta t\to 0}0\cdot\int_0^t x'(t)^2\,dt=0\,,$$

where we have applied the *Mean Value Theorem (MVT)* in the intervals $[t_i,t_{i+1}]$: $\exists\ \xi_i\in[t_i,t_{i+1}]$ such that $x(t_{i+1})-x(t_i)=x'(\xi_i)(t_{i+1}-t_i)$. From here, we further assess that

$$\int_0^T d^2x=0\quad,\quad\text{since }\int_0^T d^2x=\int_0^T 1\,dx\,dx\xleftarrow{n\to\infty}\sum_{i=1}^{n}1\cdot\left(x(t_{i+1})-x(t_i)\right)^2\overset{n\to\infty}{\to}0\,.$$

The same can be said for the covariation of two deterministic functions: $\langle x,y\rangle_T=0$, and $\int_0^T dxdy=0$.

However, if we replace these deterministic functions with <u>stochastic</u> processes – in particular with Wiener processes (standard Brownian motions), these results don't necessarily hold. Let X be a Wiener process under some probability P: $X_t\in\mathcal{N}(0,t)$, with $E[X_t]=0$, $E\left[X_t^2\right]=t$, $E[\Delta X_t]=0$, and

$E\left[\Delta^2 X_t\right] = \Delta t$ (or, equivalently, $\Delta X_t \in \mathcal{N}\left(0, \Delta t\right)$). Unlike the deterministic case, the quadratic variation of X is not zero anymore:

$$\int_0^T d^2 X \xleftarrow{n\to\infty} \sum_{i=1}^{n} E\left[\left(X\left(t_{i+1}\right) - X\left(t_i\right)\right)^2\right] = \sum_{i=1}^{n} E\left[\Delta^2 X\right] = \sum_{i=1}^{n} \Delta t = T - 0 = T$$

Furthermore,

$$\int_0^T d^2 X = T = \int_0^T dt \quad \Rightarrow \quad d^2 X = dt\,.$$ *This distorts the derivative chain rule.*

Recall Taylor's formula for a function f, of a *deterministic argument* x (more precisely, of argument $x(t)$, or x_t for short):

$$\exists\ \xi \in \left[x, x + \Delta x\right] \text{ s.a. } f\left(x + \Delta x\right) = f\left(x\right) + f_x'\left(x\right)\Delta x + \frac{1}{2} f_{xx}''\left(\xi\right)\ \Delta^2 x\,,$$ (notation: $f_{xx}''\left(\xi\right) = \frac{d^2 f}{dx^2}\left(\xi\right)$)

from where we infer that $\frac{\Delta f}{\Delta t} = f_x'\left(x\right)\frac{\Delta x}{\Delta t} + \frac{1}{2} f_{xx}''\left(\xi\right)\ \frac{\Delta^2 x}{\Delta t}$. Now note that $\frac{\Delta^2 x}{\Delta t} = \Delta x \cdot \frac{\Delta x}{\Delta t}$, and in the limits, $\Delta x \to 0$ and $\frac{\Delta x}{\Delta t} \to x'$ (a finite value), hence $\frac{\Delta^2 x}{\Delta t} \to \frac{d^2 x}{dt} = 0$. Moreover, $\xi \to x$ and $f_{xx}''\left(\xi\right) \to f_{xx}''\left(x\right)$. Is now apparent that, in the limits, the second term of the Taylor expansion vanishes, which leads us to the well-known <u>deterministic</u> chain rule $\frac{df}{dt} = f_x'\left(x\right)\frac{dx}{dt}$.

In the stochastic case, f is a function of a stochastic argument X_t, and Taylor formula looks similarly:

$\frac{\Delta f}{\Delta t} = f_x'\left(X\right)\frac{\Delta X}{\Delta t} + \frac{1}{2} f_{xx}''\left(\xi\right)\ \frac{\Delta^2 X}{\Delta t}$. However, in the limits, $\frac{\Delta^2 X}{\Delta t} \to \frac{d^2 X}{dt} = \frac{dt}{dt} = 1$, which leads to the *stochastic chain rule:*

$$\frac{df}{dt} = f_x'\left(X\right)\frac{dX}{dt} + \frac{1}{2} f_{xx}''\left(X\right) \quad ,$$

which is in fact Itô Formula: $\boxed{df = f_x'\left(X\right)dX + \frac{1}{2} f_{xx}''\left(X\right)dt}$, or in integral form,

$$f\left(X\left(b\right)\right) - f\left(X\left(a\right)\right) = \int_a^b f_x'\left(X\right)dX + \frac{1}{2}\int_a^b f_{xx}''\left(X\right)dt\,.$$

Compare these two derived formulas with the deterministic formulas: $df = f_x'(x)dx$, or in integral form, $f(x(b)) - f(x(a)) = \int_a^b f_x'(x)dx$ - that is, the *fundamental theorem of (deterministic) calculus*.

Note. Itô's formula is sometimes found in the following form: $df = f_x'(X)dX + \frac{1}{2} f_{xx}''(X)d\langle X\rangle_t$, where we replaced dt by $d\langle X\rangle_t$: $d\langle X\rangle_t = E\left[d^2X \mid \mathcal{F}_t\right] = dt$.

As an application, let's verify that the solution of the SDE $dP_t = \mu P_t dt + \sigma P_t dx_t$ (stock price dynamics – introduced later) is given by $P_t = P_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma x_t}$. Denoting $f(P_t) = \ln(P_t)$, we have:

$$d(\ln(P_t)) \overset{It\hat{o}}{=} f'(P_t)dP_t + \frac{1}{2} f''(P_t)d\langle P\rangle_t \underset{\text{only } P^2\sigma^2 dx_t^2 \neq 0}{\overset{d\langle P\rangle_t = E\left[d^2P|\mathcal{F}_t\right]}{=}} f'(P)dP + \frac{1}{2} f''(P)P^2\sigma^2 dt =$$

$$\overset{f''(P_t) = -\frac{1}{P_t^2}}{=} \frac{1}{P_t}(\mu P_t dt + \sigma P_t dx_t) - \frac{1}{2}\sigma^2 dt = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dx_t \ .$$

Then, integrating both sides, we obtain $\ln(P_t) - \ln(P_0) = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma x_t$, which is equivalent to $P_t = P_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma x_t}$. This solution will be used in the subsequent sections.

2.3 Doléans-Dade Exponential

"Research is what I'm doing when I don't know what I'm doing."
Wernher Von Braun, 1912 – 1977

Before diving into details, recall that the quadratic variation of a Brownian motion, as well as covariation of two Brownian motions are not necessarily zero. Here we consider Brownian motions with some arbitrary diffusion. If X and Y are Brownian motions, $X_t \in \mathcal{N}\left(0,\sigma_X^2 t\right)$, and $Y_t \in \mathcal{N}\left(0,\sigma_Y^2 t\right)$, with

$$E\left[X_t\right]=0\ ,\ E\left[X_t^2\right]=\sigma_X^2 t\,, \qquad E\left[Y_t\right]=0\ ,\ E\left[Y_t^2\right]=\sigma_Y^2 t$$

$$E\left[\Delta X_t\right]=0 \text{ and } E\left[\Delta^2 X_t\right]=\sigma_X^2\Delta t \qquad E\left[\Delta Y_t\right]=0 \text{ and } E\left[\Delta^2 Y_t\right]=\sigma_Y^2\Delta t$$

and, in addition,

$$E\left[X_t Y_t\right]=\rho_{XY}\sigma_X\sigma_Y t \qquad E\left[\Delta X_t \Delta Y_t\right]=\rho_{XY}\sigma_X\sigma_Y\Delta t\,,$$

then in a time-interval $\left[0,T\right]$ we have:

$$\left\langle X\right\rangle_T=\lim_{\|\Delta\|\to 0}\sum_{\Delta}E\left[\left(X_{t_{i+1}}-X_{t_i}\right)^2\right]=\lim_{\|\Delta\|\to 0}\sum_{\Delta}E\left[\Delta^2 X_t\right]=\lim_{\|\Delta\|\to 0}\sum_{\Delta}\sigma_X^2\Delta t=\sigma_X^2 T$$

$$\left\langle X,Y\right\rangle_T=\lim_{\|\Delta\|\to 0}\sum_{\Delta}E\left[\left(X_{t_{i+1}}-X_{t_i}\right)\left(Y_{t_{i+1}}-Y_{t_i}\right)\right]=\lim_{\|\Delta\|\to 0}\sum_{\Delta}E\left[\Delta X_t\Delta Y_t\right]=\rho_{XY}\sigma_X\sigma_Y\cdot\lim_{\|\Delta\|\to 0}\sum_{\Delta}\Delta t=\rho_{XY}\sigma_X\sigma_Y T$$

and finally,

$$d\left\langle X\right\rangle_t=E\left[d^2X\mid\mathcal{F}_t\right]=\sigma_X^2 dt\,, \text{ and } d\left\langle X,Y\right\rangle_t=E\left[dX\,dY\mid\mathcal{F}_t\right]=\rho_{XY}\sigma_X\sigma_Y dt\,.$$

Consider M_t^P a standard Brownian motion under a probability measure P, $M_t^P\in\mathcal{N}_P\left(0,t\right)$, and the stochastic differential equation $dZ_t=Z_t dM_t^P$, with initial condition $Z_0=1$. To solve this SDE, we apply Itô's Lemma to the function $f\left(Z_t\right)=\ln\left(Z_t\right)$:

$$df\left(Z_t\right)\overset{\text{Itô}}{=}f'\left(Z_t\right)dZ_t+\frac{1}{2}f''\left(Z_t\right)d\left\langle Z\right\rangle_t=\frac{dZ_t}{Z_t}+\frac{1}{2}\left(-\frac{1}{Z_t^2}\right)d\left\langle Z\right\rangle_t=dM_t^P-\frac{1}{2}\cdot\frac{d\left\langle Z\right\rangle_t}{Z_t^2}=$$

$$=dM_t^P-\frac{1}{2}\cdot\frac{E\left[d^2Z\mid\mathcal{F}_t\right]}{Z_t^2}\overset{dZ_t=Z_t dM_t^P}{=}dM_t^P-\frac{1}{2}\cdot\frac{Z_t^2E\left[d^2M_t^P\mid\mathcal{F}_t\right]}{Z_t^2}=dM_t^P-\frac{1}{2}\cdot d\left\langle M_t^P\right\rangle.$$

Integrating and removing $f\left(Z_t\right)=\ln\left(Z_t\right)$ by exponentiation, we obtain $Z_t=e^{M_t^P-\frac{1}{2}\cdot\left\langle M_t^P\right\rangle}$ (recall that $Z_0=1$). This expression is called *Doléans-Dade exponential*, sometimes denoted by

$$\boxed{\varepsilon\left(M^P\right)=\exp\left[M_t^P-\frac{1}{2}\left\langle M^P\right\rangle_t\right]} \quad ,$$

and is the solution of the SDE $dZ_t=Z_t dM_t^P$, with M_t standard BM and initial condition $Z_0=1$. This solution can be extended to more general underlying processes M_t. In particular, if M_t is a martingale with initial value M_0, the solution obviously becomes $\varepsilon\left(M^P\right)=\exp\left[M_t^P-M_0^P-\frac{1}{2}\left\langle M^P\right\rangle_t\right]$.

Later in this essay we will require the solution of a similar process Y_t whose dynamics is given by $dY_t=Y_t\sigma_Y dM_t^P$, with M_t^P standard Brownian motion and initial value Y_0. Note the addition of the volatility term σ_Y - for the sake of practice, let's derive the solution of this SDE as well. We apply again Itô's Lemma to $\ln\left(Y_t\right)$:

$$d\ln\left(Y_t\right)\overset{\text{Itô}}{=}\frac{dY_t}{Z_t}+\frac{1}{2}\left(-\frac{1}{Y_t^2}\right)d\left\langle Y\right\rangle_t=\sigma_Z dM_t^P-\frac{1}{2}\cdot\frac{d\left\langle Y\right\rangle_t}{Y_t^2}=\sigma_Z dM_t^P-\frac{1}{2}\cdot\frac{E\left[d^2Y\mid\mathcal{F}_t\right]}{Y_t^2}=$$

$$\overset{dY_t=\sigma_Y Y_t dM_t^P}{=}\sigma_Y dM_t^P-\frac{1}{2}\cdot\frac{\sigma_Y^2 Y_t^2 E\left[d^2M^P\mid\mathcal{F}_t\right]}{Y_t^2}=\sigma_Y dM_t^P-\frac{1}{2}\cdot\sigma_Y^2 d\left\langle M_t^P\right\rangle$$

Integrating, we obtain $\ln\left(Y_t\right)-\ln\left(Y_0\right)=\sigma_Y M_t^P-\frac{1}{2}\cdot\sigma_Y^2\left\langle M_t^P\right\rangle$, and by exponentiation, we reach the solution

$$\boxed{Y_t=Y_0 e^{\sigma_Y M_t^P-\frac{\sigma_Y^2}{2}t}},$$

where we used that $\left\langle M_t^P\right\rangle=t$. The importance of this process is due to the fact that it occurs frequently as a consequence of the martingale representation theorem. This solution will be used explicitly in Section 4.4, when dealing with the convexity adjustment for LIBOR in arrears.

3 Part B: Girsanov Theorem

"Few people have the imagination for reality."
Wolfgang von Goethe, 1749 – 1832

3.1 Part I : Girsanov Theorem for Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and Z be a non-negative random variable satisfying $E^P[Z]=1$. Then the function $Q:\mathcal{F} \to [0,1]$, given by

$$Q(A)=\int_A Z(\omega)dP(\omega) \qquad \textit{(Lebesgue integral)}$$

is a probability measure under which <u>any</u> random variable X on $(\Omega, \mathcal{F}, P)$ satisfies

$$E^Q[X]=E^P[XZ].$$

Note that $Q(A)$ is a conditional expectation under probability P: $Q(A)=E^P\left[X|A\right]$. The condition $E^P[Z]=1$ ensures that Q is indeed a probability measure:

$$Q(\Omega)=\int_\Omega Z(\omega)dP(\omega)=E^P[Z]=1.$$

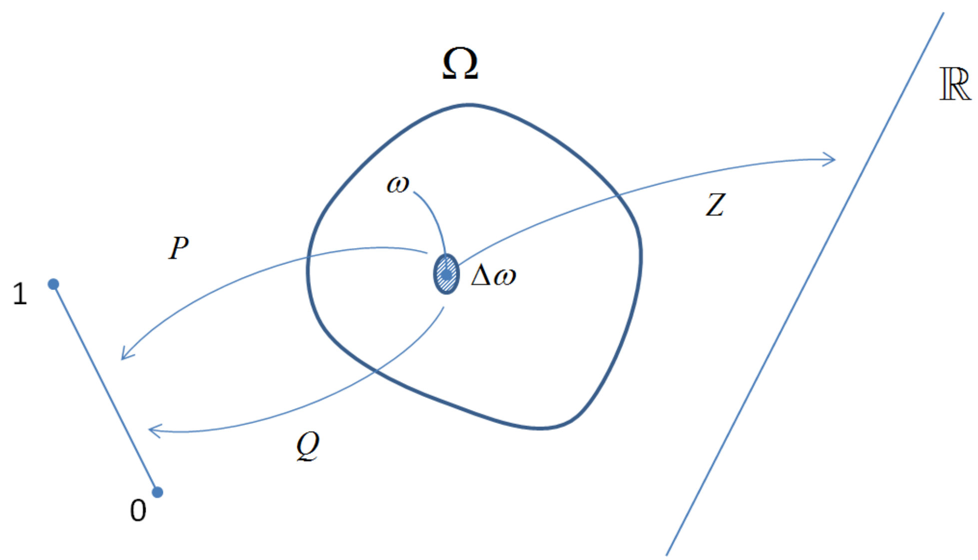
The random variable Z is called the *<u>Radon-Nikodym</u> derivative* of Q with respect to P and is formally denoted by $Z=\frac{dQ}{dP}$. If $P(Z>0)=1$, then P and Q agree on the null set (they are equivalent). Conversely, if P and Q agree on the null set, the Radon-Nikodym derivative exists and <u>is unique</u>.

<u>Radon-Nikodym derivative as a Probability Weight Function</u>

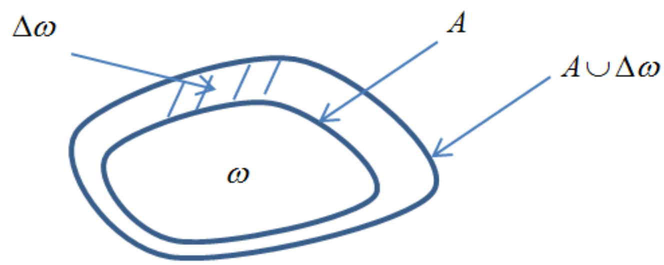
Let $(\Omega, \mathcal{F}, P)$ be a probability space. We want to define a mechanism that "tweaks" P, that is, assigns more chance to some events, and less to some others. In other words, we want to define a simple probability transformation, based on some "weights" that can be applied to P, to "tweak" the chance of events occurring in $\mathcal{F}$. That transformation will have the effect of changing the expectation of random variables, to our liking.

We seek a probability weight function on $(\Omega, \mathcal{F}, P)$, as some function $Z : \Omega \to \mathbb{R}$ that can be "applied" to P. This weight function cannot be directly applied to $P(\omega)$, for some/any outcome $\omega \in \Omega$, since $P(\omega) = 0$ most of the time (e.g., in the continuum case). It must be applied to the events in $\mathcal{F}$. One way to define such weight function is to first define it on infinitesimal events in $\mathcal{F}$ (we will define $P(d\omega)$).

In the figure below, we depict: (1) the probability space $(\Omega, \mathcal{F}, P)$, (2) the real line equipped with its corresponding Borel σ-algebra, (3) the sought-after probability weight function Z, and (4) the hypothetical probability Q obtained from P by applying the weights given by Z. In this picture, we denote by ω any outcome in Ω and by $\Delta\omega$ an infinitesimal event in $\mathcal{F}$ that contains ω, that is, an infinitesimal neighborhood of ω.



The chance of a "small" event $\Delta\omega \in \mathcal{F}$ occurring is given by $P(\Delta\omega)$. The weight that we seek to apply to $P(\Delta\omega)$ is $Z(\omega)$, that is, the new probability of $\Delta\omega$ occurring should become $Q(\Delta\omega) = Z(\omega) \cdot P(\Delta\omega)$, or better put, $Z(\omega) = \dfrac{Q(\Delta\omega)}{P(\Delta\omega)}$. Yet, this happens at infinitesimal level. Note that, having both P and Q measures - that is, satisfying the additive property - we have $Q(\Delta\omega) = \Delta Q(\omega)$ and $P(\Delta\omega) = \Delta P(\omega)$. Indeed, the following figure shows that a measure P satisfies $P(\Delta\omega) = \Delta P(\omega)$, due to its additive property:



$$P(A \cup \Delta\omega) = P(A) + P(\Delta\omega)$$
$$P(\Delta\omega) = P(A \cup \Delta\omega) - P(A) = \Delta P(\omega)$$

Here, $\Delta\omega$ is an infinitesimal neighborhood of ω. Then, in differential form, $Z(\omega) = \dfrac{dQ(\omega)}{dP(\omega)}$. At this moment it has become intuitively apparent the integral form of this expression: for any $A \in \mathcal{F}$, $Q(A) = \int_A Z(\omega)\, dP(\omega)$ - a sum of weighted probabilities across the event A. The weighting represents a transformation $Q(d\omega) = Z(\omega) \cdot P(d\omega)$, and viewing the event A as a partition $\wp_A$ into infinitesimal events $d\omega$: $A = \bigcup_{d\omega \in \wp_A} d\omega$, we can write in <u>very informal</u> terms that:

$$Q(A) = Q\left(\bigcup_{d\omega \in \wp_A} d\omega\right) = \sum_{d\omega \in \wp_A} Q(d\omega) = \sum_{d\omega \in \wp_A} Z(\omega) P(d\omega) = \sum_{d\omega \in \wp_A} Z(\omega) dP(\omega) = \int_A Z(\omega) dP(\omega) \ .$$

It makes sense to demand Z be compatible with the σ-algebra structures in Ω and $\mathbb{R}$ (that is, $\mathcal{F}$ and the Borel σ-algebra in $\mathbb{R}$, respectively). In other words, Z must be $\mathcal{F}$-measurable, i.e., a random variable!

We have built a framework around this probability transformation, however we haven't fully defined it yet. One question left unanswered is how can we ensure that Q is indeed a probability measure, in particular, that $Q(\Omega) = 1$. This happens only if $\int_\Omega Z(\omega) dP(\omega) = E^P[Z] = 1$ - hence the condition in Girsanov Theorem. To resume our findings:

<u>Definition</u>. A *probability weight function* on $(\Omega, \mathcal{F}, P)$, is a random variable $Z : \Omega \to \mathbb{R}$, that satisfies $E^P[Z] = 1$. It represents the probability transformation $P(A) \underset{Z-transformation}{\overset{weighting}{\rightarrow}} Q(A) = \int_A Z(\omega) dP(\omega)$, which has the effect of changing expectations according to $E^Q[X] = E^P[XZ]$, for all random variables X. Our probability weight function Z is exactly the Radon-Nikodym derivative.

This first part of Girsanov Theorem simply gives the mechanism of weighting a probability measure into a new probability measure, and provides the translation between expectations under these two probabilities: $E^Q[X] = E^P[XZ]$, for any random variable X . Indeed,

$$E^Q[X] = \int_\Omega X(\omega)\,dQ(\omega) = \int_\Omega X(\omega)\left[Z(\omega)\,dP(\omega)\right] = \int_\Omega \left[X(\omega)Z(\omega)\right]dP(\omega) = E^P[XZ].$$

This result is often used in practice, e.g., for the purpose of turning non-martingale processes into martingales (by eliminating drift, that is, reducing expectation to zero). The method follows a reverse direction: we usually start with a process $X_\bullet$ which is not a martingale (has drift) under a certain working probability measure P and we want to perform a change of measure (P changed into Q), so that, under the new measure, the process becomes a martingale. We proceed by guessing a process $Z_\bullet$ for which $E^P[X_\bullet Z_\bullet] = 0$ (zero drift). If we are lucky and we find such process, verifying the additional condition $E^P[Z_\bullet] = 1$, the process $X_\bullet$ under the probability measure Q defined in Girsanov Theorem will be a martingale: $E^Q[X_\bullet] = E^P[X_\bullet Z_\bullet] = 0$.

Note. The above description is stated in very general terms, hence the ambiguous notation $"X_\bullet"$ *for a process. In the next section (Part II Girsanov Theorem for Processes) we state more rigorously how the change of probability measure is actually applied to processes. Here we have just given the "algorithm" that is followed for eliminating a process' drift, in generic terms.*

In the following, we abuse terminology, and say that a random variable has "drift" if its expectation is not zero.

Example 1 - a discrete case (coin toss)

Consider the experiment of tossing a coin with sides given by T (tails) and H (heads), and assume an unbiased coin: the probabilities of tails and heads are equal. The probability space (Ω, F, P) is given by the table to the left, below:

Ω	P	X
H	0.5	3
T	0.5	5

Z	Q
0.5	0.25
1.5	0.75

This is a game of coin tossing with payoff represented by X . Ω, P and X are given, with $E^P[X] = 0.5\times 3 + 0.5\times 5 = 4$ hence the random variable X , viewed as the payoff of a game of chance, has a drift: the game is unfair. In the table to the right, the random variable Z is chosen such that $E^P[Z] = 0.5\times 0.5 + 0.5\times 1.5 = 1$. The Lebesgue integral simplifies for this discrete case into

$$Q(H) = Z(H)\times P(H) = 0.5\times 0.5 = 0.25\text{, and } Q(T) = Z(T)\times P(T) = 1.5\times 0.5 = 0.75\,.$$

The coin under Q (essentially, we have two coins: a P-coin and a Q-coin) exhibits a bias towards T (tails): T is 3 times more probable than H under Q.

We have $E^Q\left[X\right]=E^P\left[XZ\right]=0.5\times\left(3\cdot 0.5\right)+0.5\times\left(5\cdot 1.5\right)=0.75+3.75=4.5$. Note that we can compute $E^Q\left[X\right]$ directly as $E^Q\left[X\right]=0.25\times 3+0.75\times 5=4.5$. Here, we note that the new coin (Q-coin) provides a new probability Q, under which, the random variable has a different drift (larger actually!). More importantly, note that the random variable X does no change – only the probability changes. The Q–coin is obtain from the P–coin (which has no bias: H and T have equal probabilities) by, say, adjusting the weights of the two sides, H and T, so that T becomes more probable. This weighting is provided by Z : we tripled the chance of T occurring. The same random variable X (payoff) will now have a different expectation – the game is even more unfair now.

<u>Example 2 - another discrete case (coin toss)</u> : – eliminating "drift" (finding Z such that $E^Q\left[X\right]=0$)

We now show how we can turn an unfair game of chance into a fair one. Consider now the P-coin and random variable X as in the table to the left:

Ω	P	X
H	0.5	-1
T	0.5	2

Z	Q
a	a/2
b	b/2

We have $E^P\left[X\right]=0.5\times(-1)+0.5\times 2=0.5$: the game drifts the "profit" towards the tails (T). We seek another probability Q, hence a weighting function Z that transforms P into Q (with the weights $Z(H)=a$, and $Z(T)=b$), under which $E^P\left[XZ\right]=E^Q\left[X\right]=0$. Z must satisfy $E^P\left[Z\right]=1$, therefore, right away, we have to enforce the condition $0.5\times a+0.5\times b=1$. By definition, $Q(A)=\int_A Z(w)\,dP(w)$, (the continuous case form), hence $Q(H)=0.5\times a$ and $Q(T)=0.5\times b$, and since we want $E^Q\left[X\right]=E^P\left[XZ\right]=0$, we must have $0.5\times a\times(-1)+0.5\times b\times(2)=0$. Since a and b are probability weights, they must be positive; and since Q is a probability ($Q(H)\in[0,1]$ and $Q(T)\in[0,1]$), we infer that $a\in[0,2]$ and $b\in[0,2]$. We can assemble all these conditions into a system:

$$\begin{cases} 0.5\times a+0.5\times b=1 \\ 0.5\times(-1)\times a+0.5\times 2\times b=0 \\ a\in[0,2], \quad b\in[0,2] \end{cases}$$

which leads to $a=\frac{4}{3}$, $b=\frac{2}{3}$, $Q(H)=\frac{2}{3}$, $Q(T)=\frac{1}{3}$. Under probability measure Q, $E^Q[X]=0$, as Q assigns a greater probability to H, thus shifting the mean towards 0. Indeed, the weight function Z assigns a chance for H occurring greater than for T occurring: $a>b$. Using the original payoff random variable X and the new Q-coin, one can finally play a fair game. The main idea in this exercise is that we have transformed an unfair game into a fair one by only "tweaking" probabilities (the payoff given by random variable X has not been changed).

3.2 Part II : Girsanov Theorem for Processes

"The essence of mathematics lies in its freedom."
Georg Cantor, 1845 – 1918

Of our interest here are processes Y that are defined based on some given standard Brownian motions, such as $Y_t = \mu dt + \sigma dW_t^P$, with W^P standard Brownian motion under some measure P, hence $Y_t \in \mathcal{N}_P\left(\mu t, \sigma^2 t\right)$. The technique of expressing the dynamics of the process Y (as opposed to a random variable) under a different measure Q, is yet again called a change of measure, and is subject to the second part of Girsanov Theorem. Since under measure Q, W^P will most likely cease to be a martingale, the change of measure method will essentially replace W^P with another process W^Q, which is a standard Brownian motion under Q, and we seek to further express Y in terms of W^Q.

Let P and Q be probability measures on the measurable space $\left(\Omega, \mathcal{F}_\infty\right)$, W^P be a standard Brownian motion under measure P adapted to the filtration $\{\mathcal{F}_t\}_{t\in T}$, and assume that Radon-Nikodym derivative $\frac{dQ}{dP}$ exists and is represented by the process $Z(t) = E^P\left[\frac{dQ}{dP} \mid \mathcal{F}_t\right] = \varepsilon\left(M_t\right)$, that is, by a Doléans-Dade exponential , with M_t being a martingale adapted to the filtration $\{\mathcal{F}_t\}_{t\in T}$. Then the process W^Q defined as

$$\boxed{W_t^Q = W_t^P - \left\langle W^P, M\right\rangle_t}$$

is a standard Brownian motion under Q (hence a martingale). This is, essentially, Girsanov Theorem for stochastic processes. Note that Y_t can readily be expressed as $Y_t = \mu dt + \sigma\left(W_t^Q + \left\langle W^P, M\right\rangle_t\right)$, which represents the process' dynamics under measure Q.

Notation wise, observe carefully that the derivative $\frac{dQ}{dP}$ in itself is "timeless" – it connects the probability spaces $\left(\Omega, \mathcal{F}_\infty, P\right)$ and $\left(\Omega, \mathcal{F}_\infty, Q\right)$; yet, in order to perform a process transformation, we need to operate at filtration level: we need a representation of the derivative in $\mathcal{F}_t$, which is expressed naturally by a conditional expectation. If we "freeze" the processes at time t, we can see the processes as plain random variables in the corresponding probability spaces $\left(\Omega, \mathcal{F}_t, P/\mathcal{F}_t\right)$ and $\left(\Omega, \mathcal{F}_t, Q/\mathcal{F}_t\right)$ respectively. Everything must happen in $\mathcal{F}_t$, which becomes the σ-algebra that defines measurable sets in Ω. In the terminology of Girsanov Theorem Part I, $Z=\frac{dQ}{dP}$ is a random variable (probability weight function) in the probability space $(\Omega,\mathcal{F}_\infty,P)$, and $Z(t)=E^P\left[\frac{dQ}{dP}\,|\,\mathcal{F}_t\right]$provides average weights over the measurable sets in $\mathcal{F}_t$: we need only the average (in Lebesgue sense) values of Z on those sets. In this new world, $Z(t)$ is a *Radon-Nikodym process*.

Girsanov Theorem, as stated above, provides the means of rewriting an SDE of a process in a given measure P into an SDE in the equivalent measure Q , by a change of Brownian motion: simply take the SDE written in measure P and replace W_t^P with $W_t^Q+\langle W^P,M\rangle_t$: the new SDE will have its dynamics under Q: $Y_t=\mu dt+\sigma\left(W_t^Q+\langle W^P,M\rangle_t\right)$, as already stated.

Change of Expectation

Equally useful is to be able to convert conditional expectations from a probability space into the other. For a given "plain" random variable X, we already know this conversion from Part I of Girsanov Theorem: $E^Q[X]=E^P[XZ]$, where $Z=\frac{dQ}{dP}$ is the Radon-Nikodym derivative. In what follows, we will adapt this formula to processes.

Consider again a process Y_t that is based on some Brownian motion W_t^P in $(\Omega,\mathcal{F},P)$, and let $\{\mathcal{F}_t\}_{t\in T}$ be the filtration induced by W_t^P (see Section 2.1.3 for the definition of process-induced filtration). Let $Z=\frac{dQ}{dP}$ be the *Radon-Nikodym derivative* connecting the equivalent probability measures P and Q. Denote the following *Radon-Nikodym process*: $Z(t)=E^P\left[\frac{dQ}{dP}\middle|\mathcal{F}_t\right]$ (sometime called the *Radon-Nikodym density process*). Here we use a legitimate conditional expectation indeed, and we seek to find the connection between Z and $Z(t)$ (or Z_t, for short). To start with, Z_t is a martingale under P :

$$E^P[Z_t\,|\,\mathcal{F}_s]=E^P\left[E^P[Z\,|\,\mathcal{F}_t]\middle|\mathcal{F}_s\right]\overset{\text{tower property}}{=}E^P\left[Z\middle|\mathcal{F}_s\right]=Z_s \text{ , and } Z_0=E^P[Z]=1 \text{ .}$$

We know already from "Part I: Girsanov Theorem" that, since Y_t is an $\mathcal{F}_t$-measurable random variable, we must have $E^Q\left[Y_t\middle|\mathcal{F}_t\right]=E^P\left[Y_tZ_t\middle|\mathcal{F}_t\right]$ (note that everything happens in the measurable space $(\Omega,\mathcal{F}_t)$). Now, let $s\le t$ be two time points, which implies that $\mathcal{F}_s\subseteq\mathcal{F}_t$. For any event $A\in\mathcal{F}_s$, let's track the following derivation:

$$\int_A \frac{1}{Z_s} E^P\left[Y_t Z_t \middle| \mathcal{F}_s\right] dQ \overset{E^Q[1_A \bullet]=\int_A \bullet\, dQ}{=} E^Q\left[1_A \cdot \frac{1}{Z_s} E^P\left[Y_t Z_t \middle| \mathcal{F}_s\right]\right] \overset{E^Q[1_A \bullet]=E^P[1_A Z_s \bullet]}{=} E^P\left[1_A \cdot E^P\left[Y_t Z_t \middle| \mathcal{F}_s\right]\right] =$$

$$= E^P\left[E^P\left[1_A \cdot Y_t Z_t \middle| \mathcal{F}_s\right]\right] \underset{property}{\overset{tower}{=}} E^P\left[1_A Z_t Y_t\right] \overset{E^P[1_A Z_t \bullet]=E^Q[1_A \bullet]}{=} E^Q\left[1_A Y_t\right] \overset{E^Q[1_A \bullet]=\int_A \bullet\, dQ}{=} \int_A Y_t\, dQ$$

Since this derivation works for all $A \in \mathcal{F}_s$, this simply shows (using the definition of conditional expectation in Section 2.1.2) that $E^Q\left[Y_t \middle| \mathcal{F}_s\right]$ is precisely $\frac{1}{Z_s} E^P\left[Y_t Z_t \middle| \mathcal{F}_s\right]$. More conveniently, we write

$$\boxed{E^Q\left[Y_t \middle| \mathcal{F}_s\right] = E^P\left[Y_t \left(\frac{Z_t}{Z_s}\right) \middle| \mathcal{F}_s\right]}.$$

This relation, that translates the conditional expectation from one measure to the other, will be used in Section 4.3. Compare the two versions of the theorem:

random variable X, $Z = \frac{dQ}{dP}$	stochastic process $\left\{Y_t\right\}_t$, $Z_t = E^P\left[\frac{dQ}{dP} \mid \mathcal{F}_t\right]$
$E^Q\left[X\right] = E^P\left[XZ\right]$	$E^Q\left[Y_t \middle| \mathcal{F}_s\right] = E^P\left[Y_t \frac{Z_t}{Z_s} \mid \mathcal{F}_s\right]$

Finally, we stress again that $Z_0 = E^P\left[\frac{dQ}{dP} \mid \mathcal{F}_0\right] = E^P\left[\frac{dQ}{dP}\right] = E^P\left[Z\right] = 1$, by the property (or, requirement) of Z (Section 3.1).

Before diving into examples, is worth emphasizing the two available techniques, derived from Girsanov Theorem and Radon-Nikodym derivative:

- When we want to express an SDE of a process into a different new measure, we can perform an equivalent change of Brownian motion (a formal substitution, essentially).

- When we have two numéraires available, we can express conditional expectations from a measure to the other using the Radon-Nikodym process. This will be shown in Section 4.3.

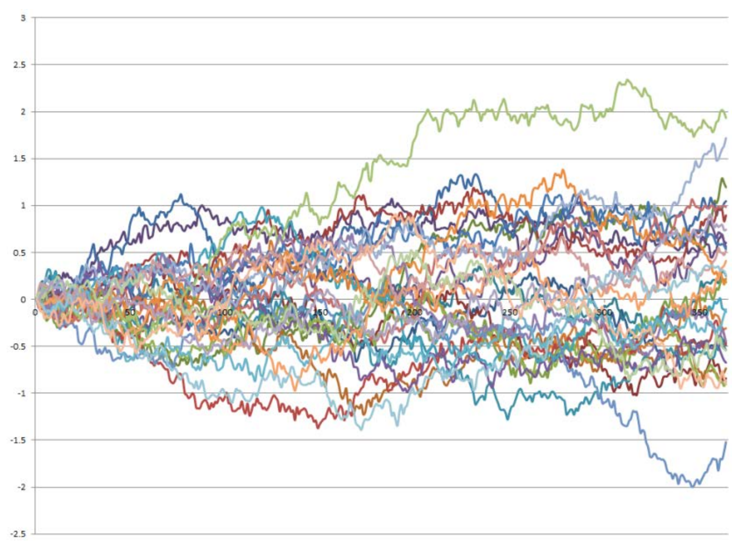
This is a summary of the constructs that we have presented so far:

Radon-Nikodym construct	what is used for
$Z = \dfrac{dQ}{dP}$	the actual R-N <u>derivative</u>: a random variable that gives the weighting transformation function between the measures P and Q
$Z(t) = E^P\left[\dfrac{dQ}{dP} \middle| \mathcal{F}_t\right]$	the R-N density <u>process</u>: a process that is used to transform a Brownian motion from a measure to the other, therefore to rewrite an SDE from one measure to the other
$\dfrac{Z(t)}{Z(s)}$	the <u>change of expectation factor</u>: a process that is used to transform conditional expectations from a measure to the other – will be used heavily in Section 4.3 and onward

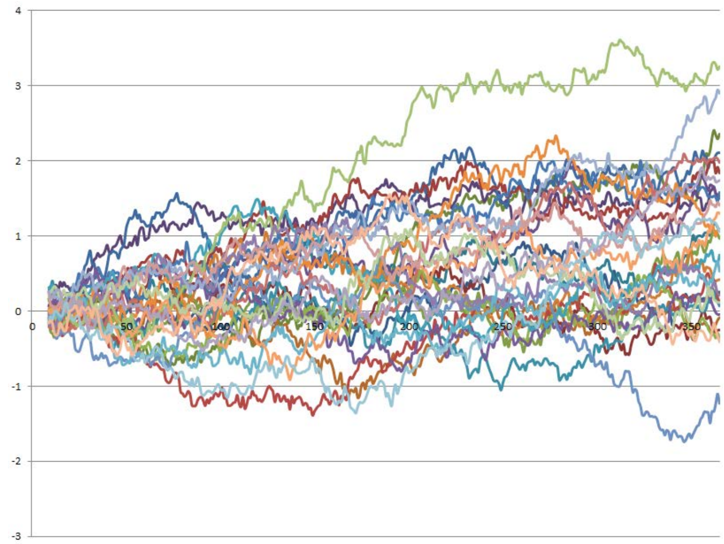
We sometimes denote $\dfrac{dQ}{dP} / \mathcal{F}_t = Z(t)$, for brevity. (note the different notations : " / " vs. "|")

<u>Example – change of measure for eliminating drift</u>

Let's start with a standard Brownian motion $W_t^P \in \mathcal{N}_P(0,t)$ under a probability measure P, and adapted to a filtration $\{\mathcal{F}_t\}_t$ (one can very well consider the natural filtration induced by the given Brownian motion) . These are 30 simulated evolution paths of W^P, showing no drift, as expected:



For exemplification, let us now construct a drifty process $Y_t = \mu t + \sigma W_t^P \in \mathcal{N}_P(\mu t, \sigma^2 t)$, that is, with drift μ and diffusion σ; in other words, $E^P[Y_t] = \mu t$ and $Var^P[Y_t] = E^P[Y_t^2] - E^P[Y_t]^2 = \sigma^2 t$. Incidentally, note the terminology: the drift is not an expectation, but rather the rate of change in expectation, and the diffusion is not a standard deviation, but rather the square root of the rate of change in variance. These are 30 simulated paths for the evolution of Y_t , for $\mu = 0.8$ and $\sigma = 1.25$ (the drift is visually apparent):



We aim at applying a change of measure from P into a new probability Q , such that Y_t becomes driftless (a martingale) under Q : $E^Q\left[Y_t\right] = 0$, yet with the same diffusion as under P: $Var^Q\left[Y\right]_t = \sigma^2 t$.

Denote by U_t the process $U_t = \frac{1}{\sigma}Y_t \in \mathcal{N}_P\left(\frac{\mu}{\sigma}t, t\right)$. In what follows, we will remove the drift of U_t, then reconstruct Y_t. For that, we follow Girsanov's statement backwards: we aim at expressing U_t as a standard Brownian motion under a probability Q, that is, by Girsanov's formula, we want to express U_t in the following form:

$U_t = W_t^P - \left\langle W_t^P, M_t^P \right\rangle$ - and now, we solve this equation for M_t^P.

We ask the question: what process M_t^P satisfies $\underbrace{\frac{\mu}{\sigma}t + W_t^P}_{U_t} = W_t^P - \left\langle W_t^P, M_t^P \right\rangle$? One can easily see that the process $M_t^P = -\frac{\mu}{\sigma}W_t^P$ verifies the expression $\left\langle W_t^P, M_t^P \right\rangle = \left\langle W_t^P, -\frac{\mu}{\sigma}W_t^P \right\rangle = -\frac{\mu}{\sigma}\left\langle W_t^P \right\rangle = -\frac{\mu}{\sigma}t$, hence $U_t = W_t^P - \left\langle W_t^P, M_t^P \right\rangle$. Therefore $M_t^P = -\frac{\mu}{\sigma}W_t^P$ is our solution. We use M_t^P to define the change of measure: $Z(t) = \varepsilon(M_t)$.

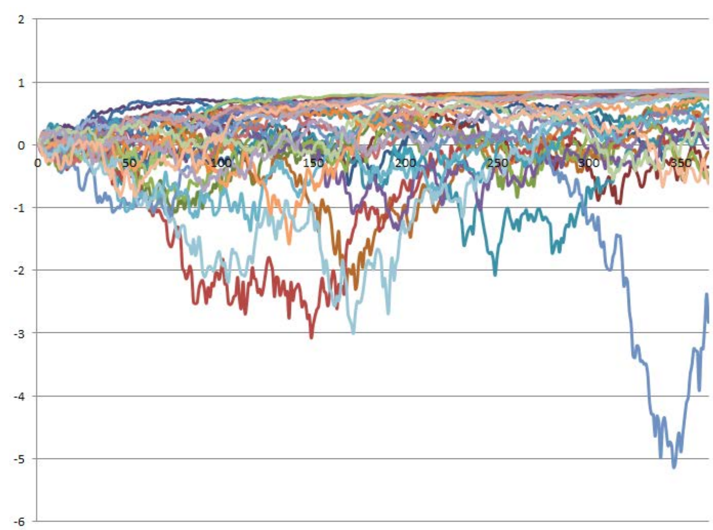
Let Z_t be a process adapted to filtration $\{\mathcal{F}_t\}_t$ and solution to the SDE $dZ_t = Z_t dM_t^P$, with $Z_0 = 1$. Then Z_t is given by the following Doléans-Dade exponential:

$$Z_t = \varepsilon\left(M^P\right) = \exp\left[M_t^P - M_0^P - \frac{1}{2}\left\langle M^P\right\rangle_t\right] = \exp\left[-\frac{\mu}{\sigma}W_t^P - \frac{1}{2}\cdot\left(-\frac{\mu}{\sigma}\right)^2\left\langle W^P\right\rangle_t\right] =$$

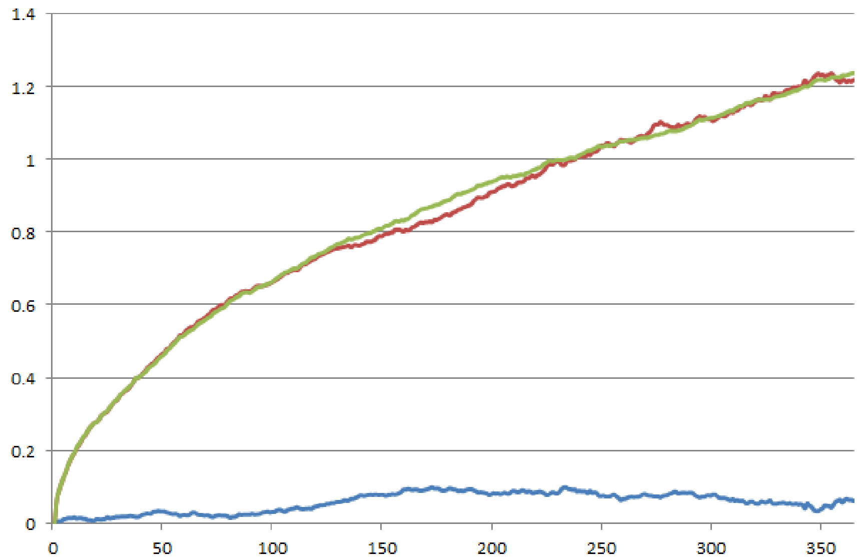
$$= \exp\left[-\frac{\mu}{\sigma}W_t^P - \frac{\mu^2}{2\sigma^2}\cdot t\right] \quad .$$

Then, Girsanov Theorem (part II) says that, under the probability measure Q, with $\frac{dQ}{dP}/\mathcal{F}_t = \varepsilon\left(M^P\right) = Z_t$, the process $W_t^P - \left\langle W^P, M\right\rangle_t$ is a martingale. And that process $W_t^P - \left\langle W^P, M\right\rangle_t$ is precisely U_t; in other words, $U_t \in \mathcal{N}_Q\left(0,t\right)$, and further, $Y_t = \sigma U_t$ is a martingale with diffusion σ.

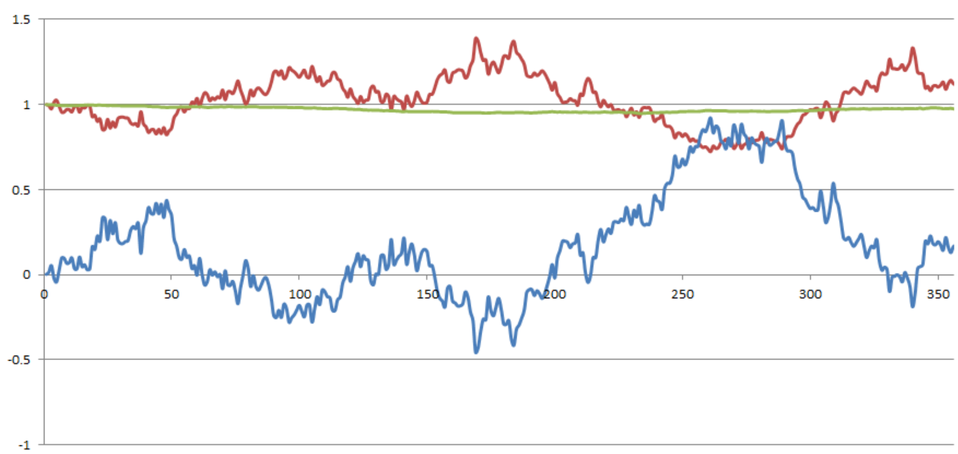
Let's test this finding. The first part of Girsanov Theorem says that $E^Q\left[Y_t\right] = E^P\left[Y_t Z_t\right]$, and we expect that, simulating $Y_t Z_t$ under P, its terminal expected value must be zero. These are the first 30 paths in the simulation of $Y_t Z_t$ under P, or equivalently, of Y_t under Q:



The expectation of $Y_t Z_t$ over 1000 paths should be (and indeed is) close to zero, at each simulated time t (guaranteed by Girsanov Theorem). The standard deviation of Y_t under Q is given by $Var^Q\left[Y_t\right] = E^Q\left[Y_t^2\right] - E^Q\left[Y_t\right]^2 = E^P\left[Y_t^2 Z_t^2\right] - E^P\left[Y_t Z_t\right]^2$, and should be precisely $Var^P\left[Y_t\right] = \sigma^2 t$, as the change of measure doesn't change variance. These facts are reflected in the following graph, where $E^Q\left[Y_t\right]=E^P\left[Y_tZ_t\right]$ is plotted in **BLUE**, $Var^P\left[Y_t\right]$ is plotted in **GREEN**, and $Var^Q\left[Y_t\right]$ is plotted in **RED**:



Finally, as stated in the first part of Girsanov Theorem , the expectation of Z_t under probability measure P should be 1, for all t. The following path plots $E^P\left[Z_t\right]$ in **GREEN**, one path for Y_t simulated under P in **BLUE**, and the corresponding path for Z_t in **RED**:



Indeed, $E^P\left[Z_t\right]$ is 1 across the timeline. One can also note that Z_t provides more weight to negative values of Y_t and less weight for positive values. This is meant to drag the expectation $E^P\left[Y_t\right]=\mu t$ down to zero.

Let's verify that indeed $E^P\left[Z_t\right]=1, \forall t$. In general, to compute the moments of a random variable $Z_t = e^{X_t}$, with X_t Gaussian, we use the moment-generating function for Gaussian random variables: $M_{X_t}(u) = E\left[e^{uX_t}\right] = e^{uE[X_t]+\frac{1}{2}Var[X_t]\cdot u^2}$ *(note the notation: $M_{X_t}(u)$ has no connection with the process M_t^P previously discussed)*. For $u=1$ and $Z_t = e^{X_t}$, this gives us the two moments:

$$E\left[Z_t\right] = e^{E[X_t]+\frac{1}{2}Var[X_t]}$$

$$Var\left[Z_t\right] = E\left[Z_t^2\right] - E\left[Z_t\right]^2 = E\left[e^{2X_t}\right] - E\left[e^{X_t}\right]^2 = e^{2E[X_t]+Var[X_t]}\cdot\left(e^{Var[X_t]}-1\right).$$

Substituting $X_t = -\frac{\mu}{\sigma}W_t^P - \frac{\mu^2}{2\sigma^2}\cdot t$, we obtain

$$E^P\left[Z_t\right] = E^P\left[\exp\left[-\frac{\mu}{\sigma}W_t^P - \frac{\mu^2}{2\sigma^2}\cdot t\right]\right] = \exp\left[-\frac{\mu^2}{2\sigma^2}\cdot t + \frac{1}{2}\cdot\left(-\frac{\mu}{\sigma}\right)^2 t\right] = e^0 = 1 \quad ,$$

as observed in our experiment. Note incidentally that $Var\left[Z_t\right] = e^{\frac{\mu^2}{\sigma^2}t} - 1$.

Let's recapitulate the journey so far. The following are the processes that we constructed, in their order of appearance:

process	distribution	descripiton
W_t^P	$\mathcal{N}_P(0,t)$	standard BM under P
$Y_t = \mu t + \sigma W_t^P$	$\mathcal{N}_P\left(\mu t, \sigma^2 t\right)$	process with drift under P and driftless under $Q : Y_t \in \mathcal{N}_Q\left(0,\sigma^2 t\right)$
$U_t = \frac{1}{\sigma}Y_t = \frac{\mu}{\sigma}t + W_t^P$	$\mathcal{N}_P\left(\frac{\mu}{\sigma}t, t\right)$	auxiliary process that becomes a standard BM under Q : $U_t \in \mathcal{N}_Q(0,t)$
$M_t^P = -\frac{\mu}{\sigma}W_t^P$	$\mathcal{N}_P\left(0, \frac{\mu^2}{\sigma^2}t\right)$	solution to the equation $U_t = W_t^P - \left\langle W_t^P, M_t^P\right\rangle$
$Z_t = \exp\left[-\frac{\mu}{\sigma}W_t^P - \frac{\mu^2}{2\sigma^2}\cdot t\right]$	$\mathcal{N}_P\left(1, e^{\frac{\mu^2}{\sigma^2}t}-1\right)$	solution to $dZ_t = Z_t dM_t^P$ with $Z_0 = 1$ ($\frac{dQ}{dP}/\mathcal{F}_t$: Radon-Nikodym derivative)

We proved that indeed, $E^Q\left[Y_t\right] = E^P\left[Y_t Z_t\right] = 0$ and $Var^Q\left[Y_t\right] = Var^P\left[Y_t\right] = \sigma^2 t$. Yet, note that we have only checked experimentally the two moments under probability P (by analyzing the process $Y_t Z_t$ under probability P), as we haven't created yet a framework for simulating Y_t under probability Q .

Therefore, a natural question arises, that is, how does the dynamic of Y_t looks like under Q? Or, equivalently, how can we simulate Y_t itself under Q? This will be the subject of the next section.

3.3 From Change of Measure to Change of Random Variable

"Since we cannot change reality, let us change the eyes which see reality."

Nikos Kazantzakis, 1883 – 1957, in Report to Greco

This section can be viewed as a prelude to Section 4.2, for the technique used here will be revisited later, in the context of the C-M-G[1] Theorem. We aim at showing how we can implement a change of measure in a concrete simulation. Although the presentation involves plain random variables, it can be easily rewritten in terms of processes by adding a time dimension to it. After all, to generate a standard Brownian motion in practice, we essentially draw random numbers from a standard normal distribution, and apply a time scale: $W(t)=\sqrt{t}\cdot\phi$, or $dW(t)=\sqrt{dt}\cdot\phi$, where ϕ is a random draw from $\mathcal{N}(0,1)$.

The title of this section may seem to contradict an observation we made repeatedly in the previous two sections, that a change of measure is <u>not</u> a change of random variable. In Section 3.1 we already noted that when performing a change of measure, we only apply weights to probabilities (thus, probabilities change), and all the other structures, including random variables, remain the same. This observation still holds strong; yet, in practice, the probability measure is hidden somewhere in the internals of random number generators (RNG), and we simply cannot tinker with it. Therefore, in practice, we do not perform changes of measure per se, but rather emulate changes of measure by variable substitutions (known as a change of variable). The take-home idea is that we merely <u>mimic the effects</u> of probability measure change by artificially changing random variables - a proxy for what we plan to achieve.

Before diving into the matter, we digress briefly for a refresher on computer number representation, and RNGs. We will use the discretized representation of real numbers to achieve some intuition on the connection between random number generators and our established framework. Let's start with the fact that a computer is a <u>finite</u> state system which cannot represent the mathematical continuum. For example, let's write down a computer view of the well-known constant π:

$$\pi_{computer}=3.14159265358979323846\ .$$

Obviously, π has an infinite decimal expansion. Then, what does the above representation stand for, really? We may very well think of $\pi_{computer}$ that represents one of the following:

$$[3.14159265358979323846, 3.14159265358979323847) \quad - \quad \text{a right-interval}$$

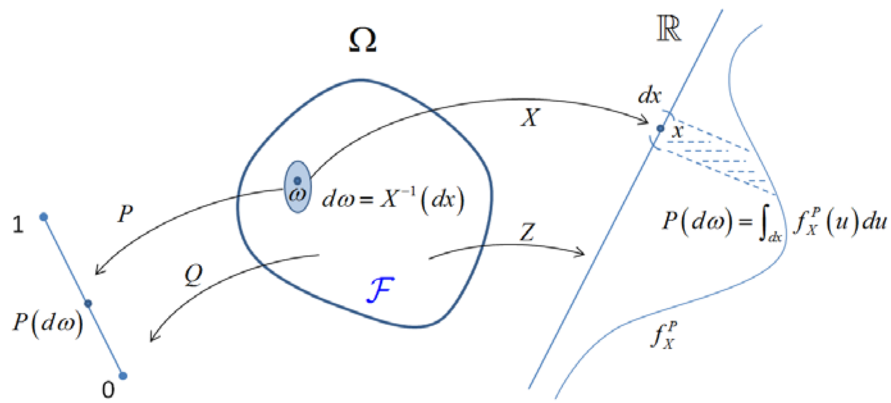
$$(3.14159265358979323845, 3.14159265358979323846] \quad - \quad \text{a left-interval}$$

$$[3.141592653589793238455, 3.141592653589793238465) \quad - \quad \text{a centered interval, etc.}$$

[1] Cameron-Martin-Girsanov

The key point here is to, sometimes, imagine a computer-represented number as an infinitesimal real-number interval. *In other words, a computer-represented number* x *will sometimes be viewed as* dx*.*

We are now ready to gain some intuition on what a RNG means in our theoretical framework. Consider the following familiar figure:



Here, we can readily identify the probability space $(\Omega, \mathcal{F}, P)$ and the real random variable $X : \Omega \to \mathbb{R}$. For an infinitesimal open interval $dx \subseteq \mathbb{R}$ (by a stretch of imagination, we view dx as a Borel set), we know by definition that $d\omega = X^{-1}(dx) \in \mathcal{F}$, hence we can reason about $P(d\omega)$: we can very well say that the probability that X will fall within dx is $P(d\omega)$. Note the probability density function f_X^P as well: $f_X^P(x)$ can be understood as the likelihood that X will fall in dx (or, in the immediate neighborhood of x) too – is an instantaneous probability concentration. Please keep in mind that a computer does not distinguish between x and dx.

In this framework, how do we interpret a function call to a RNG from a computer simulator? First, regardless of the RNG implementation we must specify the distribution from which we are drawing random numbers. In this brief discussion, let's assume we have a RNG that returns random draws in $\mathcal{N}(0,1)$, that is, standard normally distributed numbers. The only parameter we provide to a RNG call is the distribution function and we seek to fit this idea in our framework. Let's imagine that when the RNG is called, it taps into some source of entropy and randomly materializes an event $d\omega$ with probability $P(d\omega)$. Note that we have just assumed that the probability P is intimately linked to the RNG that we use! Then, the RNG will return a value x representing $X(d\omega)$: recall that we don't distinguish between dx and x, in that, somewhere along the line, the event $d\omega \in \mathcal{F}$ has been represented by a particular outcome $\omega \in d\omega$, for computer discretization's sake. In other words, the RNG will return the value $x = X(\omega)$. It is the responsibility of the RNG to make sure that the probability P will provide X with a standard normal distribution, in $\mathcal{N}(0,1)$.

This gedanken-experiment gave us an important insight: we cannot really tell the RNG what probability to use in the number generation process, and not even how to change its built-in probability (or, source of entropy): the probability P is hidden in the RNG and is inaccessible. Then, how can we perform a change of probability measure when we cannot access P ?

It turns out that there is a way: by emulating the outcome of the measure change by a random variable change – described in the following. Assume that indeed the probability P behind the RNG confers a standard normal distribution to X : $X \in \mathcal{N}_P(0,1)$ - the subscript indicates under which probability the distribution is taken from. Then, we know that the density function of X is

$$f_X^P(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$ - note the zero mean, and unit variance.

Say we want to transform probability P into another probability Q, where the transformation is governed by the following simple Radom-Nikodym derivative (or weight function, in light of Section 3.1):

$$Z(\omega) = \frac{dQ}{dP}(\omega) = e^{-\gamma X(\omega) - \frac{1}{2}\gamma^2}$$, for some given drift γ .

Indeed, this is a legitimate Radom-Nikodym derivative :

$$E^P[Z] = e^{E\left[-\gamma X(\omega) - \frac{1}{2}\gamma^2\right] + \frac{1}{2}Var\left[-\gamma X(\omega) - \frac{1}{2}\gamma^2\right]} = e^{-\frac{1}{2}\gamma^2 + \frac{1}{2}\gamma^2} = 1 .$$

To roundup the circumstances of our problem:

We have available a RNG for a random variable X , that provides random draws from a standard normal distribution f_X^P, under the probability measure P . We cannot change the RNG (or P , for that matter). Yet, we want to perform simulations under the new probability Q ; and we are given the theoretical transformation from P into Q . We know the Radon-Nikodym derivative $\frac{dQ}{dP}$; in other words, we are given a weight function that allows us to obtain probabilities in Q from probabilities in P , by multiplying the probabilities in P by some weights.

Let's rewrite the pdf of X as follows:

$$f_X^P(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+\gamma)^2}}_{f_X^Q(x)} \cdot e^{\gamma x + \frac{1}{2}\gamma^2} .$$

Note that the first factor (above) looks like a normal distribution itself. We claim that indeed, the density function (denoted by f_X^Q above) is precisely the pdf for X under probability Q ! To see this, write

$$Z(\omega)=\frac{dQ}{dP}(\omega)=\frac{Q(d\omega)}{P(d\omega)}=\frac{F_X^Q(dx)}{F_X^P(dx)}\underset{dx=X(d\omega)}{\overset{x=X(\omega)}{\approx}}\frac{dF_X^Q(x)}{dF_X^P(x)}=\frac{f_X^Q(x)}{f_X^P(x)},$$

where F denotes the CDF function, and

$$\frac{f_X^Q(x)}{f_X^P(x)}=\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x+\gamma)^2}\right)/\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)=e^{-\frac{x^2}{2}-x\gamma-\frac{\gamma^2}{2}+\frac{x^2}{2}}=e^{-\gamma x-\frac{1}{2}\gamma^2}$$, which confirms Z .

In other words, we have found that the random variable X has the density $f_X^Q(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x+\gamma)^2}$ under probability measure Q : $X\in\mathcal{N}_Q(-\gamma,1)$. This still doesn't help much, as we don't have a RNG available that is driven by Q. We must rewrite everything that happens under Q in terms of P. For that, define a new variable

$$Y=X+\gamma\ .$$

What do we know about Y? For sure, $Y\in\mathcal{N}_P(\gamma,1)$, since it is obtained from the P- standard normally distributed X, by adding the constant, γ. What is the pdf of Y under Q? Since $X=Y-\gamma$, we have that

$$f_Y^Q(y)=f_X^Q(y-\gamma)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\underbrace{y^2}_{(x+\gamma)^2}}\quad,$$

showing that $Y\in\mathcal{N}_Q(0,1)$ - that is, Y is standard normally distributed under Q . This is remarkable, as it allows us to transform any simulation under P into a simulation under Q by merely replacing X by $Y-\gamma$, and using our RNG for drawing standard normally distributed numbers for the random variable Y. The RNG has become just an instrument to draw standard normal values, which we now associate to Y instead of X; and wherever we used X in the P- probability space, we now use $Y-\gamma$. After this substitution, and by all practical purposes, we can assume we are in the Q- probability space. X in P has been metamorphosed into $Y-\gamma$ in Q This is a tabulation of our findings:

	P		Q	
X	$\mathcal{N}_P(0,1)$	$f_X^P(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$\mathcal{N}_Q(-\gamma,1)$	$f_X^Q(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x+\gamma)^2}{2}}$
Y	$\mathcal{N}_P(\gamma,1)$	$f_Y^P(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-\gamma)^2}{2}}$	$\mathcal{N}_Q(0,1)$	$f_Y^Q(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$

What we couldn't achieve by a proper change of measure, we achieved by a change of random variable.

Example – change of measure for eliminating drift revisited (see end of Section 3.2)

Let's understand better what this section has provided us with. It gave us a method to change some probability measure P into a probability measure Q, transition governed by the Radon-Nikodym derivative $Z(\omega) = \frac{dQ}{dP}(\omega) = e^{-\gamma X(\omega) - \frac{1}{2}\gamma^2}$, so that any process depending on the given random variable $X \in \mathcal{N}_P(0,1)$ as source of randomness, would evolved as if under the probability measure Q once we perform the substitution $Y = X + \gamma$. Let's rename the processes $X = \phi^P$ and $Y = \phi^Q = \phi^P + \gamma$, to avoid collusion with the notations in Section 3.2.

Recall the exercise at the end of the previous section, where we started with a drifty process under probability P, $Y_t = \mu t + \sigma W_t^P \in \mathcal{N}_P(\mu t, \sigma^2 t)$, with W_t^P standard BM, and found the Radon-Nikodym derivative for transition to a probability Q, under which Y_t becomes driftless. We checked empirically that $E^Q[Y_t] = 0$, by checking that $E^P[Y_t Z_t] = 0$, and that $E^Q[Y_t] = \sigma^2 t$ by checking that $E^P[Y_t^2 Z_t^2] - E^P[Y_t Z_t]^2 = \sigma^2 t$. However, we stopped short of simulating Y_t under Q , for lack of means to tweak the random number generator to generate events under Q. This section has provided the framework for simulating Y_t under Q, as follows.

Let's bring back the attention to $Y_t = \mu t + \sigma W_t^P$. First, we iterate the observation that W_t^P can be written as $W_t^P = \sqrt{t} \cdot \phi_t^P$, with $\phi_t^P \in \mathcal{N}_P(0,1)$; hence $Y_t = \mu t + \sigma\sqrt{t} \cdot \phi_t^P$. To simulate W_t^P amounts to drawing samples ϕ_t^P from $\mathcal{N}(0,1)$ and scaling these samples by $\sqrt{t}$. Recall the transformation in Section 3.2 governed by the Radon-Nikodym derivative $Z_t = \exp\left[-\frac{\mu}{\sigma} W_t^P - \frac{\mu^2}{2\sigma^2} \cdot t\right]$, which under the notations in this section translates to $Z(\omega) = \frac{dQ}{dP}(\omega) = e^{-\gamma \cdot \phi^P(\omega) - \frac{1}{2}\gamma^2}$: this equates to $\gamma_t = \frac{\mu}{\sigma}\sqrt{t}$. We know that $\phi_t^P \in \mathcal{N}_P(0,1)$, and that ϕ_t^P behaves under probability measure P like $\phi_t^Q = \phi_t^P + \gamma$ under probability Q. Then the dynamic of Y_t under Q must follow the equation $Y_t = \mu t + \sigma\sqrt{t} \cdot \left(\phi_t^Q - \gamma_t\right) = \mu t + \sigma\sqrt{t} \cdot \left(\phi_t^Q - \frac{\mu}{\sigma}\sqrt{t}\right) = \sigma\sqrt{t} \cdot \phi_t^Q$. This confirms what we expected: under Q , Y_t becomes driftless and furthermore $Var[Y_t] = \sigma^2 t$.

Now, this doesn't seem remarkable per se, as it only confirms what we have already expecting (it eliminates the drift and leaves the variance unchanged). The remarkable aspect is that we can change <u>all</u> processes in the P-space, applying the substitution $\phi_t^P = \phi_t^Q - \gamma$, so that we move the entire framework in the Q-space. The change of measure was accomplished essentially by performing a change of Brownian motion: $W_t^P = W_t^Q - \frac{\mu}{\sigma} t$. To simulate Y_t under the Q-measure we just sample from the standard normal distribution $\phi_t^Q \in \mathcal{N}_Q(0,1)$, and set $Y_t = \sigma\sqrt{t} \cdot \phi_t^Q$. All processes depending on ϕ_t^P under the probability P will have to replace the draws $\phi_t^P \in \mathcal{N}_P(0,1)$ by $\phi_t^Q = \phi_t^P + \frac{\mu}{\sigma}\sqrt{t}$, in order to evolve under Q . Note that the random number generator used in simulations is impartial to the probability space: it only knows how to sample from the generic standard normal distribution $\mathcal{N}(0,1)$. It is to our latitude to keep track on the change of random variable and modify all processes accordingly.

Example – change of measure in equity simulation

This matter will be dealt with in more detail in Section 4.2; here we just see how that theory fits into the change of random variable technique. But before, let's recapitulate what we found so far.

We have a transformation from probability P to probability Q, governed by the Radom-Nikodym derivative $Z(\omega) = \frac{dQ}{dP}(\omega) = e^{-\gamma \cdot \phi^P(\omega) - \frac{1}{2}\gamma^2}$. In order to change the dynamics of a process X_t^P evolving under P based on a standard Brownian motion $W_t^P = \sqrt{t}\phi^P$ (with $\phi^P \in \mathcal{N}_P(0,1)$), we simply have to replace $\phi^P = \phi^Q - \gamma$, with $\phi^Q \in \mathcal{N}_Q(0,1)$. This is equivalent to saying that we replace W_t^P by the Q Brownian motion $W_t^Q = \sqrt{t}\phi^Q = \sqrt{t}\left(\phi^P + \gamma\right) = W_t^P + \gamma\sqrt{t}$. Note that differentiating, we also have $dW_t^Q = dW_t^P + \gamma_{dt}\sqrt{dt}$. Here we emphasized that γ corresponds to a dt increment (or scaling) : γ_{dt} .

We start now with the evolution of some equity price governed by the SDE $dX_t^P = \mu X_t^P dt + \sigma X_t^P dW_t^P$. In Section 4.2 we perform a transformation from P (the "physical world" probability measure) to Q (the "risk neutral" probability measure) governed by the Radom-Nikodym derivative $Z_t = e^{-\gamma W_t^P - \frac{1}{2}\gamma^2}$, where $\gamma_t = \frac{\mu - r}{\sigma}\sqrt{t}$ (the time scale is outside of γ in the notations of Section 4.2), with the interpretation that γ is "the market price of risk". According to the above, the SDE for X_t^P changes in the Q-measure into

$$dX_t^Q = \mu X_t^Q dt + \sigma X_t^Q\left(dW_t^Q - \gamma_{dt}\sqrt{dt}\right) = \mu X_t^Q dt + \sigma X_t^Q dW_t^Q - \sigma X_t^Q \frac{\mu - r}{\sigma} dt = r X_t^Q dt + \sigma X_t^Q dW_t^Q$$

where we note that the drift vanishes, being replaced by the risk free rate r. The appeal of this transformation is that a derivative price expressed in units of numéraire is a martingale under Q.

3.4 Use Case: Quanto Adjustment for Foreign Interest Rates

"Few things are harder to put up with than a good example."
Mark Twain, 1835 – 1910

Consider the following SDE describing the evolution of FX sport rates as a geometric Brownian motion under a foreign risk-neutral probability measure Q_f :

$$dX(t) = \mu_X \cdot X(t) \cdot dt + \sigma_X \cdot X(t) \cdot dW_X^{Q_f}(t) \text{ , with solution } X(t) = X_0 \cdot e^{\left(\mu_X - \frac{\sigma_X^2}{2}\right)t + \sigma_X W_X^{Q_f}} ,$$

(assuming $X(0) = X_0$) and with the following interpretation: at time t, one unit in domestic currency is worth $X(t)$ units in foreign currency (DOM/FOR exchange rate). The standard Brownian motion used to evolve X is considered under the foreign money market measure, and is denoted by $W_X^{Q_f}$.

Note. The concept of risk-neutral measure, or money-market measure, will be introduced formally, and analyzed, in Section 4. In this section we only need to know, and use, the martingale property of an asset price expressed in units of numéraire – which will be stated shortly. We assume this property know for the time being, to provide a first simple example of change of measure that doesn't use the more subtle notion of risk neutrality.

From now on, we sometimes use the index notation for time-dependence, e.g., $X_t := X(t)$, etc.. As well, assume for the time being that the domestic and foreign short rates are constant, r_d and r_f - this is merely a notation convenience, as the time-dependence of short rates plays no significant role in the derivations that immediately follow.

We first derive the drift μ_X of the FX spot rate X_t. Let B_t^f be a numéraire representing the foreign risk-free money market account (MMA). It evolves according to $\frac{dB_t^f}{B_t^f} = r_f \cdot dt$ (the drift of its return is the short rate, and $B_0^f = 1$), hence we have that $B_t^f = e^{r_f \cdot t}$. Similarly, we have $B_t^d = e^{r_d \cdot t}$ in the domestic space. The foreign MMA expressed in units of domestic numéraire is a martingale under the domestic risk-neutral measure (by the fundamental theorem of asset pricing) and is given by $\frac{B_t^f / X_t}{B_t^d}$.

This means that the process $e^{(r_f - r_d)t} X_t^{-1} = X_0^{-1} \cdot e^{(r_f - r_d) - \left(\mu_X - \frac{\sigma_X^2}{2}\right)t - \sigma_X W_X^{Q_f}}$ must be driftless; which only happens if $\mu_X = r_f - r_d$, and the above SDE for X_t under the foreign risk-neutral measure becomes

Part B: Girsanov Theorem

$$X_t = X_0 \cdot e^{\left(r_f - r_d - \frac{\sigma_X^2}{2}\right)t + \sigma_X W_X^{Q_f}}.$$

Let V_t be any tradable asset denominated in the foreign currency. By the fundamental theorem of asset pricing, we have that, under the foreign risk-neutral measure Q^f, $\frac{V_t}{B_t^f}$ is a Q^f-martingale. Similarly, under the domestic risk-neutral measure Q^d, $\frac{V_t / X_t}{B_t^d}$ is a Q^d-martingale: the foreign asset value converted to domestic currency and divided by the domestic numéraire is a martingale under the domestic risk-neutral probability measure. By martingale property we then have:

$$\begin{cases} \dfrac{V_0}{B_0^f} = E^{Q^f}\left[\dfrac{V_t}{B_t^f} \mid \mathcal{F}_0\right] := E_0^{Q^f}\left[\dfrac{V_t}{B_t^f}\right] & \text{(eq.I)} \\ \dfrac{V_0 / X_0}{B_0^d} = E_0^{Q^d}\left[\dfrac{V_t / X_t}{B_t^d}\right] & \text{(eq.II)} \end{cases}.$$

Using these two equations, we can write the following derivations:

$$\begin{cases} 1 \overset{\text{(eq.I)}}{=} E_0^{Q^f}\left[\dfrac{V_t}{B_t^f}\right] \cdot \dfrac{B_0^f}{V_0} = E_0^{Q^f}\left[\dfrac{V_t}{B_t^f} \cdot \dfrac{B_0^f}{V_0}\right] \\ 1 \overset{\text{(eq.II)}}{=} E_0^{Q^d}\left[\dfrac{V_t / X_t}{B_t^d}\right] \cdot \dfrac{B_0^f}{V_0 / X_0} = E_0^{Q^d}\left[\dfrac{V_t / X_t}{B_t^d} \cdot \dfrac{B_0^f}{V_0 / X_0}\right] \end{cases} \Rightarrow \quad E_0^{Q^f}\left[\dfrac{V_t}{B_t^f} \cdot \dfrac{B_0^f}{V_0}\right] = E_0^{Q^d}\left[\dfrac{V_t / X_t}{B_t^d} \cdot \dfrac{B_0^f}{V_0 / X_0}\right]$$

Is time to invoke the Girsanov Theorem: for any random variable Y, the equation $E^{Q_d}\left[Y\right] = E^{Q_f}\left[Y \cdot \frac{dQ^d}{dQ^f}\right]$ holds , where $\frac{dQ^d}{dQ^f}$ is the Radon-Nikodym derivative of Q^d with respect to Q^f. We already have the equation

$$E_0^{Q^f}\left[\frac{V_t}{B_t^f} \cdot \frac{B_0^f}{V_0}\right] = E_0^{Q^d}\left[\frac{V_t / X_t}{B_t^d} \cdot \frac{B_0^f}{V_0 / X_0}\right],$$

in which we must identify $\frac{dQ^d}{dQ^f}$. Applying Girsanov Theorem for $Y_t = \frac{V_t / X_t}{B_t^d} \cdot \frac{B_0^f}{V_0 / X_0}$, we must have

$$\frac{V_t}{B_t^f} \cdot \frac{B_0^f}{V_0} = \frac{V_t / X_t}{B_t^d} \cdot \frac{B_0^f}{V_0 / X_0} \cdot \left(\frac{dQ^d}{dQ^f} / \mathcal{F}_t\right)$$

which, solving for $\frac{dQ^d}{dQ^f} / \mathcal{F}_t$, leads to the expression $\frac{dQ^d}{dQ^f} / \mathcal{F}_t = \frac{B_0^f}{B_t^f} \cdot \frac{B_t^d \cdot X_t}{B_0^d \cdot X_0}$, that can further be simplified as $\frac{dQ^d}{dQ^f} / \mathcal{F}_t = \frac{e^{r_f \cdot 0}}{e^{r_f \cdot t}} \cdot \frac{e^{r_d \cdot t}}{e^{r_d \cdot 0}} \cdot \frac{X_t}{X_0} = \frac{X_t}{X_0} \cdot e^{(r_d - r_f)t}$. To summarize, the Radon-Nikodym derivative is finally given by

$$\frac{dQ^d}{dQ^f} / \mathcal{F}_t = \frac{X_t}{X_0} \cdot e^{(r_d - r_f)t} .$$

Let's now take a closer look at the SDE for the FX spot rate: $X_t = X_0 \cdot e^{\left(r_f - r_d - \frac{\sigma_X^2}{2}\right)t + \sigma_X W_X^{Q_f}}$, and rearranging, $\frac{X_t}{X_0} \cdot e^{(r_d - r_f)t} = e^{-\frac{1}{2}\sigma_X^2 t + \sigma_X W_X^{Q_f}}$. In other words,

$$\boxed{\frac{dQ^d}{dQ^f} / \mathcal{F}_t = e^{\sigma_X W_X^{Q_f} - \frac{1}{2}\sigma_X^2 t}}$$

One should carefully note that although we reached an expression for the Radon-Nikodym derivative by starting with a process X with dynamics in the foreign risk-neutral measure Q^f, $\frac{dQ^d}{dQ^f} / \mathcal{F}_t$ is independent of X : is a change of measure rather than a change of random variable, and will apply to any random variable, for that matter. Furthermore, our initial assumption that the short rates are constant doesn't impact the expression for Radon-Nikodym derivative (the rate terms vanished).

Even more interestingly, observe that the Radon-Nikodym derivative is a Doléans-Dade exponential: $\frac{dQ^d}{dQ^f} / \mathcal{F}_t = \varepsilon(M_t) = e^{M_t - \frac{1}{2}\langle M \rangle_t}$, with $M_t = \sigma_X W_X^{Q_f}$. Therefore, the second part of Girsanov Theorem tells us that, starting with **<u>any</u>** standard Brownian motion under Q^f, $W_t^{Q^f}$, the process $W_t^{Q^d} := W_t^{Q_f} - \left\langle W^{Q_f}, M \right\rangle_t$ is a standard Brownian motion under Q_d. This process is given by

$$\boxed{W_t^{Q^d} := W_t^{Q_f} - \left\langle W^{Q_f}, \sigma_X W_X^{Q_f} \right\rangle_t = W_t^{Q_f} - \rho_{f,X} \sigma_X t}$$

Let's apply these findings to the dynamics of short rates. Consider now that the short rates <u>do evolve</u> in time, based on the following SDEs, in the two currencies under consideration (domestic and foreign), based on <u>Hull-White dynamics</u>:

$$\begin{cases} dr_d(t) = \kappa_d \cdot \left(\vartheta_d(t) - r_d(t) \right) \cdot dt + \sigma_d \cdot dW_{r_d}^{Q_d}(t) \\ dr_f(t) = \kappa_f \cdot \left[\vartheta_f(t) - r_f(t) \right] \cdot dt + \sigma_f \cdot dW_{r_f}^{Q_f}(t) \end{cases}$$

where we express r_d under the domestic money market measure, and r_f under the foreign money market measure. In the above, κ_d is the rate's mean-reversion speed and ϑ_d is its mean-reversion level. As usual, by Q_d and Q_f we have denoted the domestic and foreign risk-neutral probability measures. Finally, we highlight the correlation between the standard Brownian motions involved, $W_X^{Q_f}$ and $W_{r_f}^{Q_f}$, under the foreign risk-neutral probability measure:

$$d\left\langle W_X^{Q_f}, W_{r_f}^{Q_f} \right\rangle = \rho_{f,X} dt$$

where we applied the formula $d\left\langle X, Y \right\rangle_t = E\left[dX dY \mid \mathcal{F}_t \right] = \rho_{XY} \sigma_X \sigma_Y dt$, with $\sigma_X = \sigma_Y = 1$, for the covariation of two Brownian motions, presented previously in Section 2.3.

We aim at changing the evolution of the foreign short rate into an evolution under the domestic risk-neutral measure. The SDE in foreign risk-neutral measure is

$$dr_f(t) = \kappa_f \cdot \left[\vartheta_f(t) - r_f(t) \right] \cdot dt + \sigma_f \cdot dW_{r_f}^{Q_f}(t) \quad .$$

Here, we have described the short rate based on a standard Brownian motion $W_{r_f}^{Q_f}$ under Q^f. To express the SDE in the domestic risk-free rate Q^d, we use the previous findings, and derive the transformation

$$W_t^{Q^d} = W_t^{Q_f} - \rho_{f,X} \sigma_X t ,$$

which leads to the dynamics of the short rate under the domestic measure Q^d, given by

$$dr_f(t) = \kappa_f \cdot \left[\vartheta_f(t) - r_f(t) \right] \cdot dt + \sigma_f \cdot d\left[W_t^{Q_f} - \rho_{f,X} \sigma_X t \right] =$$

$$= \kappa_f \cdot \left[\vartheta_f(t) - r_f(t) - \rho_{f,X} \sigma_X \sigma_f \right] \cdot dt + \sigma_f \cdot dW_t^{Q_d} \quad .$$

The story would not be complete if we left the exchange rate dynamics in terms of X_t , that is, DOM/FOR rate, and described under the foreign risk-neutral measure. The last step in our endeavor is to (1) express X_t under the domestic measure, and (2) replace it with the FOR/DOM rate, X_t^{-1} (the notation reflects that this is the "inverse" exchange rate).

First, let us perform a change of measure for X_t. Recall that the SDE that we started with:

$$dX(t) = (r_f - r_d) \cdot X(t) \cdot dt + \sigma_X \cdot X(t) \cdot dW_X^{Q_f}(t) \;, \quad X(t) = X_0 \cdot e^{\left(r_f - r_d - \frac{\sigma_X^2}{2}\right)t + \sigma_X W_X^{Q_f}(t)}$$

where, yet again for convenience, we drop the time-dependence of short rates, and $W_X^{Q_f}$ is a Brownian motion under the foreign measure Q_f . We aim at rewriting this SDE under the domestic measure, Q_d .

Let us apply Girsanov Theorem for the Brownian motion $W_X^{Q_f}$: the following process is a Brownian motion under Q_d :

$$W_X^{Q_d}(t) := W_X^{Q_f}(t) - \left\langle W_X^{Q_f}, M \right\rangle_t ,$$

where $M_t = \sigma_X W_X^{Q_f}$ has been found before. Therefore, $\left\langle W_X^{Q_f}, M \right\rangle_t = \sigma_X \left\langle W_X^{Q_f} \right\rangle_t = \sigma_X t$, and we rewrite the SDE under the domestic measure as

$$X(t) = X_0 \cdot e^{\left(r_f - r_d - \frac{\sigma_X^2}{2}\right)t + \sigma_X\left[W_X^{Q_d}(t) + \sigma_X t\right]} = X_0 \cdot e^{\left(r_f - r_d + \frac{\sigma_X^2}{2}\right)t + \sigma_X W_X^{Q_d}(t)} .$$

Having this done, we now apply the second transformation: replace the DOM/FOR rate X_t with the FOR/DOM rate, X_t^{-1}. We make the following observations:

1. As the notation suggests, $X_t^{-1} = \dfrac{1}{X_t}$, and is inferred from the definition of FX rates.
2. Intuitively clear, the diffusion of X_t and X_t^{-1} are equal: $\sigma_X = \sigma_{X^{-1}}$.
3. Recalling that $\rho_{f,X}$ is the correlation between X_t and the foreign short rate, we infer a correlation of opposite sign for $W_{X^{-1}}^{Q_d}$ and $W_{r_f}^{Q_d}$: $d\left\langle W_{X^{-1}}^{Q_d}, W_{r_f}^{Q_d} \right\rangle = -\rho_{f,X} dt$. For this to happen, we must have $W_{X^{-1}}^{Q_d} = -W_X^{Q_d}$.

Note that the last two observations can be directly inferred from the first one. Indeed, from

$$X(t) = X_0 \cdot e^{\left(r_f - r_d + \frac{\sigma_X^2}{2}\right)t + \sigma_X W_X^{Q_d}(t)}$$

we derive $X^{-1}(t) = X_0^{-1} \cdot e^{-\left(r_f - r_d + \frac{\sigma_X^2}{2}\right)t - \sigma_X W_X^{Q_d}(t)} = X_0^{-1} \cdot e^{\left(\mu_{X^{-1}} + \frac{\sigma_{X^{-1}}^2}{2}\right)t + \sigma_{X^{-1}} W_{X^{-1}}^{Q_d}(t)}$, and comparing the stochastic terms of the two expressions, we directly infer that $\sigma_X = \sigma_{X^{-1}}$ and $W_{X^{-1}}^{Q_d} = -W_X^{Q_d}$.

Using the above three observations, we can write the following derivation:

$$X^{-1}(t) = X^{-1}(0) \cdot e^{-\left(r_f - r_d + \frac{\sigma_X^2}{2}\right)t - \sigma_X W_X^{Qd}(t)} \Rightarrow X^{-1}(t) = X^{-1}(0) e^{\left(r_d - r_f - \frac{\sigma_X^2}{2}\right)t + \sigma_X W_{X^{-1}}^{Q_d}(t)}$$

or equivalently,

$$dX^{-1}(t) = \left(r_d - r_f\right) \cdot X^{-1}(t) \cdot dt + \sigma_X \cdot X^{-1}(t) \cdot dW_{X^{-1}}^{Q_d}(t).$$

Wrapping up, we ended up with the following system of SDEs that describes the dynamics of domestic and foreign short rates (under Hull-White dynamics), as well as exchange rate (FOR/DOM this time), all under the domestic risk-neutral measure Q^d :

$$\begin{cases} dr_d(t) = \kappa_d \cdot \left(\vartheta_d(t) - r_d(t)\right) \cdot dt + \sigma_d \cdot dW_d^{Q^d}(t) \\ dr_f(t) = \left(\kappa_f \cdot \left[\vartheta_f(t) - r_f(t)\right] - \rho_{f,X} \cdot \sigma_f \cdot \sigma_X\right) \cdot dt + \sigma_f \cdot dW_f^{Q^d}(t) \\ \dfrac{dX(t)}{X(t)} = \left(r_d(t) - r_f(t)\right) \cdot dt + \sigma_X \cdot dW_X^{Q^d}(t) \end{cases}$$

Here, we replaced the notation X^{-1} with X for brevity, keeping in mind that X is now the FOR/DOM FX rate. The term $\rho_{f,X} \cdot \sigma_f \cdot \sigma_X \cdot dt$ is called the *quanto adjustment* for the foreign short rate dynamics. The system above is well-suited for Monte-Carlo simulations.

4 Part C: Applications

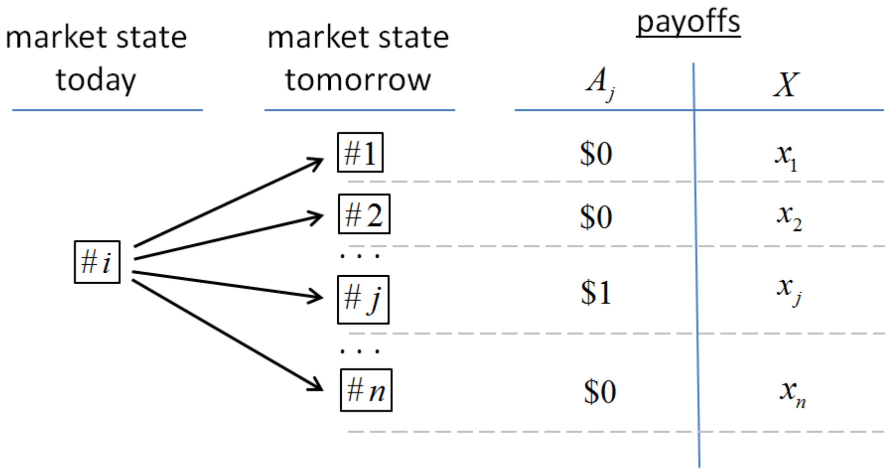
"Everything you can imagine is real."
Pablo Picasso, 1881 – 1973

4.1 Introduction: Arrow-Debrew Securities and Risk-Neutral Probability

In this section we provide an economic interpretation of so-called risk-neutral probabilities. We assume that the market is *complete*, in that every tradable asset (security) can be *replicated* by a portfolio of some *elementary securities*. A possible analogy is a vector space, say, $\mathbb{R}^n$, with a basis of unit vectors, $B=\{(1,0...0),(0,1...0),...,(0,0...1)\}$, so that every vector in $\mathbb{R}^n$ can be written as a linear combination of vectors in B. In this analogy, the vector space is our market place, the basis is our collection of elementary securities, and any vector in $\mathbb{R}^n$ is a tradable security in the market. To say that the market is complete is similar to saying that the B is a basis in $\mathbb{R}^n$. To express a vector as a linear combination of vectors in the basis is similar to saying that a tradable asset is replicated by a portfolio of elementary securities. Finally, a security that does not belong to the basis is called a *redundant security*, as it can be represented by a collection of elementary securities.

Let's construct the set of elementary securities. We start with a market, which can be found in one of n possible states. Assume that it currently is in state $\#i$ and can advance tomorrow in any of the states $\#1$, $\#2$,..., $\#n$. To start with, we assume that in our market there is no time value of money (zero risk-free interest rate). We define an Arrow(-Debrew) security A_j as a security that tomorrow will pay $\$1$ if the market advances to some state $\#j$ and will pay $\$0$ in any other market states (note the analogy of the basis vectors in $\mathbb{R}^n$). We therefore have n types of Arrow securities. If today we hold a portfolio consisting of one Arrow security of each type, the value of the portfolio (MTM) tomorrow will be precisely $\$1$: if the market will advance to a state $\#k$, the Arrow security A_k will be worth its payoff of $\$1$ and all the others will be worthless, netting to $\$1$. Since there is no time-value for the money ($\$1$ tomorrow is worth $\$1$ today), we infer that this portfolio is worth $\$1$ today, as well. Note that we don't know the price of each individual Arrow security; yet we know that the total value of all n Arrow securities is precisely $\$1$.

Any redundant security X , with arbitrary payoffs x_j in market state $\#j$, can be replicated by a portfolio of Arrow securities. Indeed, consider the following illustration:



A portfolio $\prod$ which consists of x_1 quantity of A_1, x_2 quantity of A_2, ..., x_n quantity of A_n - or, notation-wise, $\prod = \{x_1 A_1, ..., x_n A_n\}$, has the same payoff (tomorrow) as X. For example, if the market will advance to state $\# j$, all securities in $\prod$ will be worthless, except for A_j, hence $\prod$ will be worth $\$1 \cdot x_j$, precisely as the payoff of security X.

What is the value of $\prod$ today? To assess this, we need to consider the price of all arrow securities A_j: let's denote by a_j the price of A_j. Then obviously, the price of $\prod$ is $\sum_1^n a_j x_j$, and since $\prod$ replicates X, the price of X today must also be $\sum_1^n a_j x_j$. And we already know that $\sum_1^n a_j = 1$ (the price of a portfolio consisting of precisely on A_1, one A_2, ..., and one A_n, is $\$1$).

Note that we don't know the probabilities $p_1, ..., p_n$ of the market advancing in the states $\#1, ..., \#n$, respectively. However, since $\sum_1^n a_j = 1$, and since the price of $\prod$ today is $\sum_1^n a_j x_j$, we can hypothetically *imagine a world* in which security X will pay x_1 with probability a_1, x_2 with probability a_2, ..., and x_n with probability a_n. If this was the case, then in this *imaginary* world, today's <u>expected payoff</u> of $\prod$ would be precisely $\sum_1^n a_j x_j$, which is in fact the true value of X *in the real-world* today. So, for pricing purpose alone, we may very well replace the real probabilities $p_1, ..., p_n$ with the quantities $a_1, ..., a_n$, and proceed with computing the expected payoff of X under these fictitious new probabilities: we'll obtain an accurate price of X today, in the real-world.

We have defined two important notions: (1) the <u>real-world</u> (or, physical world) in which the market evolved based on some unknown <u>real probabilities</u> $p_1, ..., p_n$; and (2) an <u>imaginary-world</u>, in which the market evolves based on some <u>fictitious</u> (yet, mathematically sound) probabilities $a_1,\dots,a_n$; and we discovered that the price in real-world, today, of a real-world security, can be viewed as the expected value, today, of the security in the imaginary-world. We also defined <u>the probabilities in the imaginary-world to be the unit price of Arrow securities</u>. Nomenclature-wise, we call the imaginary-world the "*risk-neutral world*" and the fictitious probabilities the "*risk-neutral probabilities*" (that is, Arrow prices).

We have just defined risk-neutral probabilities as being the market unit price of Arrow securities. However, what is the nature of these quantities? We know that they are determined by the supply and demand in the market, driven by economic factors, such as:

- **[hedging arguments]** The preferences of the market participants with respect to holding money tomorrow in one market state versus another.

- **[time value of money]** The preference with respected to holding money today versus tomorrow.

- **[real-world risk]** The estimated probabilities that the market will actually evolve in a specific state (real-world probabilities $p_1,\dots,p_n$).

The last factor deserves special attention. It says that the real-world risk is embedded in the unit price of Arrow securities, which are nothing but our risk-neutral probabilities – this explains the terminology: for pricing a security in the risk-neutral world, we simply perform an expectation under the risk-neutral probabilities, and do not account for any risk inherent to the actual security under consideration (that risk is already accounted for, indirectly, by the risk-neutral probabilities themselves).

If we take into consideration the time value of money, we require a small tweak to the Arrow security prices, in order to still use them as (risk-neutral) probabilities.

Consider now that, in this market, the risk-free borrowing or lending money can be done at a risk-free rate r. Let's rebuild the previous framework under these circumstances. We have the same set of Arrow securities, and the same redundant security X that will pay x_1, …, x_n, based on the market state attained tomorrow.

The portfolio $\{A_1,\dots,A_n\}$ of just Arrow securities is still worth $\$1$ tomorrow, but only $\frac{1}{1+r}$ today, due to discounting at risk-free rate. . The market pays today a_j for Arrow security A_j, therefore, we have that $\sum_1^n a_j = \frac{1}{1+r}$. This shows that we cannot use $\{a_j\}_j$ as the set of risk-neutral probabilities anymore, as they don't add up to 1 – we have to tweak them. Indeed, let's define the following quantities: $\{q_j = a_j \cdot (1+r)\}_j$ - a <u>scaling</u> of $\{a_j\}_j$. We now have $\sum_1^n q_j = 1$, and we seek to use $\{q_j\}_j$ as risk-neutral probabilities.

The portfolio $\prod = \{x_1 A_1, ..., x_n A_n\}$ of fractional Arrow securities still replicates X, for, regardless of the risk-free rate, the payoff of $\prod$ tomorrow is identical to that of X. Today, the portfolio costs $\sum_1^n a_j x_j$ to buy. The expected value under $\{q_j\}_j$ of portfolio tomorrow is $\sum_1^n q_j x_j$ (we assume that X will take each value x_j with probability q_j). Then, this value discounted to today is $\frac{1}{1+r}\sum_1^n q_j x_j$, and if $\{q_j\}_j$ are indeed risk-neutral probabilities, it must match the cost of the portfolio today:

$$\sum_1^n a_j x_j \overset{??}{=} \frac{1}{1+r}\sum_1^n q_j x_j ,$$

and this is obviously true, since $q_j = a_j \cdot (1+r)$ for all $j \in \{1,..,\mathrm{n}\}$. We have found that under probabilities given by $\{q_j\}_j$ (measure Q, notation-wise), the value of security X is the discounted payoff expectation of X. In this framework, the risk neutral probabilities are the **compounded** Arrow security prices.

In practice, we do not know either the real-world probabilities, nor the risk-neutral world probabilities. Yet, we know that, for a given market, the risk-neutral probabilities are unique (Fundamental Theorem of Asset Pricing), and we know that under risk-neutral probabilities, today's price of a security must be equal to its discounted expected payoff. The last statement is equivalent to saying that the price of a security in units of numéraire is a martingale under risk-neutral probability measure.

These facts help us perform a change of measure and express a process evolving under the real-world probability P as a process evolving under the risk-neutral probability Q (without knowing the actual measures P and Q), and compute the present price of the security modeled by this process as a discounted expectation. The following section describes such a measure change. Note carefully that, since Arrow securities are abstract constructs representing a given market, the risk-neutral probabilities are independent of any particular security, and therefore, they allow the pricing of any redundant security as a discounting expectation under the risk-neutral measure.

4.2 Change of Measure using Cameron-Martin-Girsanov Theorem

"The only way to achieve the impossible is to believe it is possible."
Alice Kingsleigh in Alice Through the Looking Glass, Lewis Carroll, 1832 – 1898

Here we provide a simplistic view of *Cameron-Martin-Girsanov Theorem*, as a tool for changing the measure from a real-world probability measure to a risk-neutral probability measure. Under such change, we expect that a standard Brownian motion (BM), which is obviously driftless, would change into a BM with drift under the new measure. Our first concern is to assess this new drift.

We start with a standard BM W_t^P under a given probability P, that is, with $W_t^P \in \mathcal{N}_P(0,t)$. In the context of Girsanov Theorem, we want to study a particular change of measure, from P to some measure Q, given by the Radon-Nikodym derivative

$$\frac{dQ}{dP}/\mathcal{F}_t = \varepsilon(M_t) = e^{M_t - \frac{1}{2}\langle M\rangle_t} \text{, with } M_t = -\gamma W_t^P \text{ ,} \qquad \text{(or, } \frac{dQ}{dP}/\mathcal{F}_t = e^{-\gamma W_t^P - \frac{1}{2}\gamma^2 t} \text{)}$$

that is, given by a very simple Doléans-Dade exponential, with a parameter γ. Let's first check whether $E_P\left[\varepsilon(M_t)\right]=1$, using the moment-generating function:

$$E^P\left[\varepsilon(M_t)\right] = e^{E^P\left[M_t - \frac{1}{2}\langle M\rangle_t\right] + \frac{1}{2}Var\left[M_t - \frac{1}{2}\langle M\rangle_t\right]} = e^{E^P\left[-\gamma W_t^P - \frac{1}{2}\gamma^2 t\right] + \frac{1}{2}Var\left[-\gamma W_t^P - \frac{1}{2}\gamma^2 t\right]} = e^{-\frac{1}{2}\gamma^2 t + \frac{1}{2}\gamma^2 t} = 1 \text{ .}$$

W_t^P remains Gaussian under Q (i.e., a normally distributed random variable), with a drift that is yet to be found. We now express the moment-generating function for W_t^P under Q, that is, $M_{W_t^P}^Q(u) = E^Q\left[e^{uW_t^P}\right]$, by invoking Girsanov Theorem applied to the random variable $e^{u\cdot W_t^P}$:

$$E^Q\left[e^{u\cdot W_t^P}\right] = E^P\left[e^{u\cdot W_t^P}\cdot\frac{dQ}{dP}/\mathcal{F}_t\right] = E^P\left[e^{u\cdot W_t^P}\cdot e^{-\gamma W_t^P - \frac{1}{2}\gamma^2 t}\right] = E^P\left[e^{(u-\gamma)W_t^P - \frac{1}{2}\gamma^2 t}\right] =$$

$$= e^{-\frac{1}{2}\gamma^2 t + \frac{1}{2}(u-\gamma)^2 t} = e^{-u\cdot\gamma\cdot t + \frac{1}{2}u^2 t} \text{, which must equal } M_{W_t^P}^Q(u) = e^{uE^Q\left[W_t^P\right] + \frac{1}{2}u^2\cdot Var\left[W_t^P\right]}.$$

Since this is the moment-generating function for W_t^P, we conclude that W_t^P has the expectation $-\gamma t$ (equivalently, drift $-\gamma$) under Q : $E^Q\left[W_t^P\right] = -\gamma t$. On the other hand, Girsanov Theorem says that the process $W_t^Q = W_t^P - \left\langle W^P, M\right\rangle_t = W_t^P + \gamma t$, is a standard BM under Q. In summary,

$$\begin{cases} W_t^P \in \mathcal{N}_P(0,t) \\ W_t^P \in \mathcal{N}_Q(-\gamma t,t) \\ W_t^Q \in \mathcal{N}_P(\gamma t,t) \\ W_t^Q \in \mathcal{N}_Q(0,t) \end{cases},$$

which gives a complete picture of the nature of the two Brownian motions under each probability measure. This is a simplistic view of *Cameron-Martin-Girsanov (C-M-G) Theorem*.

We now use these findings and express the dynamics of a stock price under the risk-neutral probability measure. Under the real-world probability measure P, the stock prices evolve according to the following SDE:

$$\frac{dX_t^P}{X_t^P} = \mu dt + \sigma dW_t^P \quad , \quad \text{with } X_0^P = x_0 \ ,$$

where W_t^P is a standard BM under P, $\frac{dX_t^P}{X_t^P}$ can be viewed as an instantaneous rate of stock's returns, and μ, σ are return's drift and diffusion, respectively. Applying Itô's Lemma, the solution of this SDE is given by

$$X_t^P = x_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t^P} \ .$$

We know that there <u>exists</u> a <u>unique</u> probability measure Q (the risk-neutral measure) under which the stock price expressed in numéraire units is a martingale (Fundamental Theorem of Asset Pricing - FTAP, under the arbitrage-free assumption). More precisely, let's consider a numéraire (bank account) B_t which evolves according to $\frac{dB_t}{B_t} = rdt$ and $B_0 = 1$, that is, it accumulates interest at the risk-free rate, with solution $B_t = e^{rt}$. The random variable $\frac{X_t^P}{B_t}$ is a martingale under the risk-neutral measure Q, therefore we have $E^Q\left[\frac{X_t^Q}{B_t} \mid \mathcal{F}_0\right] = E^Q\left[e^{-rt} X_t^Q \mid \mathcal{F}_0\right] = x_0$, where by X_t^Q we denoted the process X_t evolving under Q this time (and x_0 is just a notation).

Therefore, we now must express X_t under the risk-neutral measure Q, and for doing so, we use the findings in the first part of this section: the relationship between Q and P is reflected in the following Radon-Nikodym derivative, and standard BM under Q:

$$\frac{dQ}{dP} / \mathcal{F}_t = e^{-\gamma W_t^P - \frac{1}{2}\gamma^2 t}, \quad W_t^Q = W_t^P + \gamma t \quad ,$$ for some parameter γ yet to be found.

Note. The reason why we try to apply the C-M-G transformation to our particular stock is that we are guaranteed the uniqueness of Q, therefore Q must be the measure described in the C-M-G Theorem.

This means that the random variable X_t has the following form under Q:

$$X_t^Q = x_0 e^{\left(\mu - \frac{\sigma^2}{2} - \sigma\gamma\right)t + \sigma W_t^Q},$$

and expressed in numéraire units, is a martingale under $Q: E^Q\left[e^{-rt} X_t^Q \mid \mathcal{F}_0\right] = x_0$. This translates into:

$$x_0 = E^Q\left[x_0 e^{\left(-r+\mu-\frac{\sigma^2}{2}-\sigma\gamma\right)t+\sigma W_t^Q}\right] = x_0 E^Q\left[e^{\left(-r+\mu-\frac{\sigma^2}{2}-\sigma\gamma\right)t+\sigma W_t^Q}\right] = x_0 e^{\left(-r+\mu-\frac{\sigma^2}{2}-\sigma\gamma\right)t+\frac{1}{2}\sigma^2 t} = x_0 e^{(-r+\mu-\sigma\gamma)t}$$

This can happen only if $-r + \mu - \sigma\gamma = 0$, giving the expression for parameter γ : $\boxed{\gamma = \frac{\mu - r}{\sigma}}$. We can now summarize our findings:

- The stock price under real-world measure is given by $X_t^P = x_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t^P}$.

- The stock price under risk-neutral measure is given by $X_t^Q = x_0 e^{\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t^Q}$, or in differential form, $\frac{dX_t^Q}{X_t^Q} = r \cdot dt + \sigma dW_t^Q$. Note that under probability measure Q, stock's rate of growth drifts according to the risk-free rate (or, money market rate).

- The transformation (change of measure) is given by Radon-Nikodym derivative $\frac{dQ}{dP} / \mathcal{F}_t = e^{-\frac{\mu-r}{\sigma} W_t^P - \frac{1}{2}\left(\frac{\mu-r}{\sigma}\right)^2 t}$.

- Finally, the relationship between Brownian motions under these two measures is given by $W_t^Q = W_t^P + \frac{\mu - r}{\sigma} t$. Here, $\frac{\mu - r}{\sigma}$ can be viewed as the market price of risk (*Sharpe ratio*: the asset return premium per unit of risk), and is intuitively included in the Brownian motion under the risk-neutral probability - similar to saying that is captured by the risk-neutral probability itself.

We close this discussion with an observation about the existence and uniqueness of the risk-neutral probability measure. In the above derivation, for a given security X, we notice that the transformation from P into Q is parameterized, with parameter γ_X, found to depend on the drift of stock's returns, μ_X : $\gamma_X = \frac{\mu_X - r}{\sigma_X}$ (here, the index X reflects the dependence on stock X). However, we know from the Fundamental Theorem of Asset Pricing that Q is unique, hence the transformation from P into Q must be unique, and consequently γ_X should not depend on X. This implies that the ratio $\frac{\mu_X - r}{\sigma_X}$ has a generic value, regardless of the choice of X. And this resonates with the *Capital Asset Pricing Model* (CAPM): *"the reward to risk ratio for any individual security in the market is equal to the market reward-to-risk ratio".* Here we imagine a "market portfolio" M, with its own μ_M and σ_M, and for any security X, we must have $\frac{\mu_X - r_X}{\sigma_X} = \frac{\mu_M - r}{\sigma_M}$, hence $\gamma = \frac{\mu_M - r}{\sigma_M}$ (reason for which γ is called the market price of risk), and the measure change as well as Q itself are unique indeed, and independent of any particular asset X. Under measure Q, any asset price, expressed in units of numéraire, is a martingale, and any asset's returns will evolve with a drift given by the risk-free rate r.

Furthermore, any redundant security can be priced as a discounted expectation: a replicating portfolio of martingales is a martingale itself, under the unique risk-neutral measure. Indeed, consider a redundant security X, replicated by a portfolio $\prod$ (for all t, $X(t) = \prod(t)$) consisting of n assets: $\prod(t) = \sum_1^n \lambda_i(t) S_i(t)$, where $\{\lambda_i(t)\}_i$ are adapted/predictable stochastic processes governing the allocation of assets at time t. We further require that the portfolio is self-financing; that is, the change in value of $\prod$ is due only to the change in value of its constituent assets: $d\prod(t) = \sum_1^n \lambda_i(t) dS_i(t)$, or in integral form, $\prod(T) - \prod(t) = \sum_1^n \int_t^T \lambda_i(u) dS_i(u)$. If S_i (expressed in units of numéraire) are martingales under the risk-neutral measure, the integrals on the right side are Itô integrals, which are martingales as well, hence $\prod$ is a sum of martingales, therefore a martingale itself.

Notes.

1. *In **Monte Carlo simulations**, sometimes we simulate stock/asset prices under the risk-neutral measure, and some other times we simulate them under the real/physical measure. Risk-neutral measure is suitable, and mainly used, for pricing purposes; whereas for other reasons, such as back-testing, testing investment strategies, portfolio optimizations, etc., one uses the real measure. For the latter, one must find ways to estimate the drift μ and diffusion σ of the simulated asset, e.g., using historical observations.*

2. *The main purpose of this change of measure is to price a derivative as a discounted expectation. Let's see how this works in a Monte Carlo simulation framework. If* $V\left(S\left(t\right)\right)$ *is the price of a derivative, with asset* S *as underlying, and with maturity* T*, and if we simulate* S *under the risk-neutral measure* Q*, and simply average the terminal values* $V\left(S\left(T\right)\right)$ *and discount them to the present, we obtain today's value of the derivative:*

$$V_0 = \frac{1}{B\left(T\right)} E^Q\left[V\left(S\left(T\right)\right) \mid \mathcal{F}_0\right].$$

If we perform the simple arithmetic average of all the realized payoff values at time T*, across all simulated paths, is as if we summed-up realized values multiplied by their frequencies of occurrence; that is, the discrete way of averaging values weighted by their probabilities. Thus, we agree that this sum is an expectation. Now, recalling that* $\mathcal{F}_0 = \left\{\varnothing, \Omega\right\}$*, is clear that the expectation is conditioned by filter* $\mathcal{F}_0$*. The take-home idea here is that the arithmetic average of all simulation paths in a Monte Carlo simulation is more properly described as a discrete integration over the entire domain* $\mathbb{R}$ *(or equivalently, expectation across entire* Ω *).*

4.3 Change of Numéraire – a Working Formula

"Simplicity is the ultimate sophistication."
Leonardo da Vinci, 1452 – 1519

Is no secret that, for the sake of tractability, we prefer to deal with martingales rather than arbitrary drifty processes, wherever possible. For example, we have already made the case that the payoff $V(t)$ of some financial instrument expressed in units of numéraire is a martingale under the risk-neutral probability measure. Therefore we often prefer to work with the quantity $\frac{V(t)}{B(t)}$, with $B(t)$ money market account (MMA, or MM account, or bank account) numéraire, rather than $V(t)$ alone. The reason is that, under the risk-neutral measure Q_B, and for two time points $s \le t$, we have $\frac{V(s)}{B(s)} = E^{Q_B}\left[\frac{V(t)}{B(t)} \mid \mathcal{F}_s\right]$, which gives us more leverage in deriving the pricing analytics. We will see in the next Section 4.4 that there are different choices of numéraire: say, a bank account or a bond, as convenience dictates. So far we dealt only with a measure change from the physical measure P to the risk-neutral measure Q_B associated to the MMA numéraire. Here we aim at finding a simple formula that translates between expectations under two probability measures Q_B and Q_N, associated to two different numéraires, $B(t)$ and $N(t)$.

The stage consists of a measurable space $(\Omega, \mathcal{F}_\infty)$, a filtration $\{\mathcal{F}_t\}_t$, and two probability measures Q_B and Q_N, under which $\frac{V_t}{B_t}$, and $\frac{V_t}{N_t}$ are martingales, respectively. Therefore, for two time points $s \le t$, we already know by definition that $\frac{V_s}{B_s} = E^{Q_B}\left[\frac{V_t}{B_t} \mid \mathcal{F}_s\right]$ and that $\frac{V_s}{N_s} = E^{Q_N}\left[\frac{V_t}{N_t} \mid \mathcal{F}_s\right]$. This means that $V_s = B_s E^{Q_B}\left[\frac{V_t}{B_t} \mid \mathcal{F}_s\right] = N_s E^{Q_N}\left[\frac{V_t}{N_t} \mid \mathcal{F}_s\right]$. Denoting $Y_t = \frac{V_t}{B_t}$ the martingale process under Q_B, we have:

$$E^{Q_B}\left[Y_t \mid \mathcal{F}_s\right] = \frac{N_s}{B_s} E^{Q_N}\left[\frac{V_t}{N_t} \mid \mathcal{F}_s\right] = \frac{N_s}{B_s} E^{Q_N}\left[\frac{Y_t B_t}{N_t} \mid \mathcal{F}_s\right] = E^{Q_N}\left[Y_t \frac{B_t / N_t}{B_s / N_s} \mid \mathcal{F}_s\right].$$

Now is time to recall the Radon-Nikodym (density) process $Z_t = E^{Q_N}\left[\frac{dQ_B}{dQ_N} \middle| \mathcal{F}_t\right]$, with $Z_0 = 1$, defined in Section 3.2 and connecting the expectations under Q^B and Q^N: $E^{Q_B}\left[Y_t \middle| \mathcal{F}_s\right] = E^{Q_N}\left[Y_t \frac{Z_t}{Z_s} \middle| \mathcal{F}_s\right]$.

Comparing the two expectation formulas above, we are guessing that $\frac{Z_t}{Z_s} = \frac{B_t / N_t}{B_s / N_s}$, and by variable separation and from condition $Z_0 = 1$, this only can happen if $Z_t = \frac{B_t / B_0}{N_t / N_0}$.

In summary, and generalizing, in order to convert the conditional expectation of some process Y_t from Q_N to Q_B we only have to know that

$$E^{Q_B}\left[Y_t \middle| \mathcal{F}_s\right] = E^{Q_N}\left[Y_t \frac{B_t / N_t}{B_s / N_s} \middle| \mathcal{F}_s\right]$$

where B_t and N_t are the corresponding numéraires. We will use this relation in the next section. Moreover, the Radon-Nikodym process for the change of measure from Q_N to Q_B is fully determined by the underlying numéraires N and B, and is given by

$$Z_t = \frac{B_t / B_0}{N_t / N_0}.$$

4.4 T-Forward Measure

"In mundo pressuram habetis; sed confidite, Ego vici mundum."
John, 16:33

Motivation

Let's recall the dynamics of a money market account, $\frac{d}{dt}B(t)=B(t)r(t)$, with $B(0)=1$, expressing that the account grows at rate $r(t)$, known as short (or instantaneous) rate. We have already established that the price of a security V expressed in units of B-numéraire is a martingale under the risk-neutral probability Q_B. In other words,

$$\frac{V(t)}{B(t)}=E^{Q_B}\left[\frac{V(T)}{B(T)}\,|\,\mathcal{F}_t\right], \qquad \text{or,} \qquad V(t)=E^{Q_B}\left[\frac{B(t)}{B(T)}\cdot V(T)\,|\,\mathcal{F}_t\right].$$

In the following table we refresh a few formulae, assumed known to the reader. Note the peculiar form of the MM account value for the case of a stochastic short rate: it does not require an expectation conditioned to the filter $\mathcal{F}_t$, as it depends only on the history up to time $t<T$ (or, we sometimes say, $B(t)$ is observable at time t).

short rate $r(t)$:	**stochastic**	**deterministic (assumed known)**	**constant**
MM account $B(t)$:	$e^{\int_0^t r(u)du}$	$e^{\int_0^t r(u)du}$	e^{rt}
ZC bond $P(t,T)$:	$E^{Q_B}\left[e^{-\int_t^T r(u)du}\,\|\,\mathcal{F}_t\right]=e^{-\int_t^T f(t,u)du}$	$e^{-\int_t^T r(u)du}=e^{-\int_t^T f(t,u)du}=\frac{B(t)}{B(T)}$	$e^{-r(T-t)}$
instantaneous forward rate $f(t,T)$:	$-\frac{\delta}{\delta T}\ln P(t,T)$	$r(T)$	r
forward rate $f(t,T,S)$:	$=\frac{\ln P(t,T)-\ln P(t,S)}{S-T}=\frac{1}{S-T}\int_S^T f(t,u)du$		

Note carefully that the zero coupon bond yields the risk-free rate under the risk-neutral probability Q_B, similar to a stock under the risk-neutral measure, whose returns grow at the risk-free rate.

Note also that, one may wish to think of the instantaneous forward rate $f(t,T)$ as being an expectation, $E^{Q_{--}}\left[r(T)\,|\,\mathcal{F}_t\right]$, that is, the rate paid on an instant deposit at time T, as viewed at an earlier time t. However this is not the case under the risk-neutral measure (!); yet, we will show that this *holds under the T-forward measure, defined in the following. These matters will become apparent by the end of this section.*

Finally, note that the zero coupon bond expressed in units of B-numéraire, is a martingale under Q_B:

$$\frac{P(t,T)}{B(t)} = E^{Q_B}\left[e^{-\int_0^T r(u)du} \mid \mathcal{F}_t\right] = E^{Q_B}\left[\frac{1}{B(T)} \mid \mathcal{F}_t\right] = E^{Q_B}\left[\frac{P(T,T)}{B(T)} \mid \mathcal{F}_t\right].$$

The price V of an interest rate (IR) derivative (that is, which has an IR instrument as underlying) depends, directly or otherwise, on the short rate dynamics; hence can be generically expressed under the risk neutral measure as a martingale $\frac{V\left(\underline{t},T,r\left(\underline{t}\right)\right)}{B\left(\underline{t}\right)} = E^{Q_B}\left[\frac{V\left(\underline{T},T,r\left(\underline{T}\right)\right)}{B\left(\underline{T}\right)} \mid \mathcal{F}_t\right]$, or equivalently,

$$V\left(t,T,r(t)\right) = E^{Q_B}\left[e^{-\int_t^T r(u)du} \cdot V\left(T,T,r(T)\right) \mid \mathcal{F}_t\right] ,$$

with $r(t)$ being considered stochastic, in order to capture more accurately the nature of the underlying. This expectation cannot be evaluated readily, as it has two stochastic factors. The *T-forward measure* facilitates this evaluation.

<u>Change of Numéraire: from MM Accont to ZC Bond</u>

In the following we will perform a change of measure, as a consequence of a numéraire change: replacing $B(t)$ with $P(t,T)$. That is, the MM account numéraire will be replaced with the zero-coupon bond (ZCB). The new measure Q_T will be called the *T-forward measure*; under which, the price of security V expressed in ZCB numéraire units becomes a martingale. In other words, we seek to enforce

$$\frac{V\left(t,T,r(t)\right)}{P(t,T)} = E^{Q_T}\left[\frac{V\left(T,T,r(T)\right)}{P(T,T)} \mid \mathcal{F}_t\right] = E^{Q_T}\left[V\left(T,T,r(T)\right) \mid \mathcal{F}_t\right] ,$$

where we used that $P(T,T)=1$. Recall that $P(t,T) = E^{Q_B}\left[e^{-\int_t^T r(u)du} \mid \mathcal{F}_t\right]$, giving the following expression for security V :

$$V\left(t,T,r(t)\right) = \underbrace{E^{Q_B}\left[e^{-\int_t^T r(u)du} \mid \mathcal{F}_t\right]}_{P(t,T)} \cdot E^{Q_T}\left[V\left(T,T,r(T)\right) \mid \mathcal{F}_t\right] ,$$

expression which is easier to evaluate, because of the break-down into two separate expectations. It remains to find the probability measure Q_T, or equivalently, the Radon-Nikodym derivative $\frac{dQ_T}{dQ_B} / \mathcal{F}_t$ that changes measure from Q_B to Q_T.

In the following we emphasize only the first argument of $V(t,\_,\_)$, for brevity. Since most derivations involve conditional expectations, probably more important than finding the random variable $Z = \frac{dQ_T}{dQ_B}$ is finding the Radon-Nikodym density process $Z(t) = E^{Q_B}\left[\frac{dQ_T}{dQ_B} | \mathcal{F}_t\right]$, and even more importantly, $\frac{Z(T)}{Z(t)}$: the expectation conversion factor. This process has already been expressed in Section 4.3: for these particular numéraires, we have

$$Z(t) = \frac{P(t,T)/P(0,T)}{B(t)/B(0)} \text{ and } \frac{Z(T)}{Z(t)} = \frac{P(T,T)/B(T)}{P(t,T)/B(t)} \overset{P(T,T)=1}{=} \frac{B(t)}{B(T)P(t,T)} = \frac{1}{P(t,T)} e^{-\int_t^T r(u)du} .$$

This will allow us to write, e.g., $E^{Q_T}\left[V(T) | \mathcal{F}_t\right] = E^{Q_B}\left[V(T)\frac{1}{P(t,T)} e^{-\int_t^T r(u)du} | \mathcal{F}_t\right]$, if ever needed.

We record the following formula, for the expectation conversion factor:

$$\boxed{\frac{Z(T)}{Z(t)} = \frac{1}{P(t,T)} e^{-\int_t^T r(u)du}}$$

In the above, we readily used the change of numéraire formula developed in Section 4.3. Yet, for pedagogical purpose, is worth going through the derivation once again, under the present circumstances. We start by noting that the price of V, computed under either Q_B or Q_T, must be the same:

$$V(t) \overset{\text{in } Q_B}{=} B(t) \cdot E^{Q_B}\left[\frac{V(T)}{B(T)} | \mathcal{F}_t\right] \overset{\text{in } Q_T}{=} P(t,T) \cdot E^{Q_T}\left[V(T) | \mathcal{F}_t\right]$$

leading to the equality $E^{Q_B}\left[\frac{B(t)}{B(T)} \cdot V(T) | \mathcal{F}_t\right] = E^{Q_T}\left[P(t,T) \cdot V(T) | \mathcal{F}_t\right]$. On the other hand, we have shown in Section 3.2 that, for any process X, $E^{Q_T}\left[X(T)|\mathcal{F}_t\right]=E^{Q_B}\left[X(T)\frac{Z(T)}{Z(t)}|\mathcal{F}_t\right]$.

Choosing $X(T)=P(t,T)\cdot V(T)$, we obtain $E^{Q_T}\left[X(T)\,|\,\mathcal{F}_t\right]=E^{Q_B}\left[X(T)\frac{P(T,T)/B(T)}{P(t,T)/B(t)}\,|\,\mathcal{F}_t\right]$.

Considering that necessarily $Z(0)=1$, after a variable separation and a suitable scaling, we obtain $Z(t)=E^P\left[\frac{dQ_T}{dQ_B}|\mathcal{F}_t\right]=\frac{P(t,T)/P(0,T)}{B(t)/B(0)}$, and $\frac{Z(T)}{Z(t)}=\frac{B(t)}{B(T)P(t,T)}=\frac{1}{P(t,T)}e^{-\int_t^T r(u)du}$. We have just reconstructed the same Radom-Nikodym formula.

To recap, the Radon-Nikodym derivative, just found, defines a measure change from the risk-neutral measure Q_B to a so-called *T-forward measure* Q_T, under which, the price of an IR derivative V expressed in units of ZCB numéraire is a martingale. Furthermore, we can evaluate V as

$$\boxed{V(t)=P(t,T)\cdot E^{Q_T}\left[V(T)\,|\,\mathcal{F}_t\right]}.$$

Note carefully, that the ZCB numéraire $P(t,T)$ is chosen to have its maturity coincide with the maturity of the derivative V (which recall that is denoted fully as $V(t,T,r(t))$). In some sense, the maturity T is a constant parameter – reason for which the measure is called T-forward: for every T there is an unique corresponding measure Q_T. More informally, we say that the T-forward measure is defined on the time horizon $[0,T]$.

<u>A by-Product Martingale under the T-Forward Measure: the Instantaneous Forward Rate</u>

Recall the expression that connects the instantaneous forward rate with the zero coupon bond:

$$f(t,T)=-\frac{\delta}{\delta T}\ln P(t,T)\;\;,$$

and we also know the expression for a ZC bond price, $P(t,T)=E^{Q_B}\left[e^{-\int_t^T r(u)du}\,|\,\mathcal{F}_t\right]$ - they have been mentioned in the table at the beginning of this section. Differentiating, we obtain:

$$f(t,T)=-\frac{\delta}{\delta T}\ln P(t,T)=-E^{Q_B}\left[\frac{1}{P(t,T)}\cdot\frac{\delta}{\delta T}e^{-\int_t^T r(u)du}\,|\,\mathcal{F}_t\right]=E^{Q_B}\left[\frac{1}{P(t,T)}\cdot e^{-\int_t^T r(u)du}\cdot r(T)\,|\,\mathcal{F}_t\right]$$

Using the inverse Radon-Nikodym conversion factor $\left[\frac{Z(T)}{Z(t)}\right]^{-1}$, we convert the above expectation into an expectation in T-forward measure and further obtain

$$
\begin{aligned}
f(t,T) &= E^{Q_T}\left[\frac{1}{P(t,T)}\cdot e^{-\int_t^T r(u)du}\cdot r(T)\cdot\frac{Z(t)}{Z(T)}\,|\,\mathcal{F}_t\right]= \\
&= E^{Q_T}\left[\frac{1}{P(t,T)}e^{-\int_t^T r(u)du}\cdot r(T)\cdot\frac{P(t,T)}{e^{-\int_t^T r(u)du}}\,|\,\mathcal{F}_t\right]= E^{Q_T}\left[r(T)\,|\,\mathcal{F}_t\right]= \\
&= E^{Q_T}\left[f(T,T)\,|\,\mathcal{F}_t\right]
\end{aligned}
$$

which shows that the forward rate f behaves as a martingale under Q_T, in its first argument. This resonates well with intuition: the instantaneous forward rate $f(t,T)$ can very well be written as $f(t,T,T+dt)$ - i.e., the interest rate observed at time t on an instant deposit at time T - and is what we expect at t to be the short rate at T.

<u>An Example: Vasicek short rate under the T-Forward measure</u>

The Vasicek short rate model expresses the short-rate dynamics under the risk-neutral measure Q_B by the following SDE:

$$dr(t)=a\big(b-r(t)\big)dt+\sigma dW^{Q_B}(t)\ ,$$

with the solution given by

$$r(t)=e^{-a(t-s)}r(s)+b\left(1-e^{-a(t-s)}\right)+\sigma\int_s^t e^{-a(t-u)}dW^{Q_B}(u),$$

showing that, when t approaches infinity, $r(t)\in\mathcal{N}_{Q_B}\left(b,\frac{\sigma^2}{2a}\right)$. We aim at rewriting these dynamics under the T-forward measure, for some given time horizon T. Here, a and b are the speed and level of the rate's mean-reversion, respectively.

This time, we will not seek a Doléans-Dade exponential form for the Radon-Nikodym derivative, but rather use the martingale property of the process $d\left(\frac{B(t)}{P(t,T)}\right)$ under Q_T (as $\frac{B(t)}{P(t,T)}$ is a legitimate martingale under Q_T) to directly derive the expression for W^{Q_T} (the Brownian motion under Q_T).

Consider now that the ZC bond price maturing at T evolves according to the following SDE in Q_B :

$dP(t,T) = r(t)P(t,T)dt + \tilde{\sigma}(t)P(t,T)dW^{Q_B}(t)$, with bond instantaneous volatility denoted by $\tilde{\sigma}$.

We digress briefly to justify that a ZCB price under the risk-neutral measure follows the above dynamics. We know that the ZCB price expressed in units of numéraire, $P_B(t,T) = \dfrac{P(t,T)}{B(t)}$*, is a martingale under* Q_B*. By the martingale representation theorem, we can write*

$$dP_B(t,T) = \widehat{\sigma}(t)dW^{Q_B}(t) \; ,$$

for some, possibly stochastic, function $\widehat{\sigma}(t)$*. The differential is taken over the first argument. Now define* $\tilde{\sigma}(t) = \widehat{\sigma}(t)/P_B(t,T)$*, which leads to a new SDE for the ZCB price:*

$$dP_B(t,T) = P_B(t,T)\tilde{\sigma}(t)dW^{Q_B}(t) \; .$$

Knowing that $dB(t) = r(t)B(t)dt$*, and the above, we now perform the following derivation:*

$$dP(t,T) = d\left(P_B(t,T)B(T)\right) = P_B(t,T)dB(T) + dP_B(t,T)B(T) =$$

$$= \frac{P(t,T)}{B(t)}r(t)B(t)dt + B(t)\left(\frac{P(t,T)}{B(t)}\tilde{\sigma}(t)dW^{Q_B}(t)\right) \; ,$$

leading to the sought formula:

$$dP(t,T) = r(t)P(t,T)dt + \tilde{\sigma}(t)P(t,T)dW^{Q_B}(t) \; .$$

Now, let's expand the expression for $d\left(\dfrac{B(t)}{P(t,T)}\right)$, which is a martingale under Q_T, as $d\left(\dfrac{B(t)}{P(t,T)}\right) =$

$= B(t)d\dfrac{1}{P(t,T)} + \dfrac{1}{P(t,T)}dB(t)$. Denoting $X(t) = \dfrac{1}{P(t,T)}$, and invoking Itô's Lemma, we have:

$$dX(t) = \frac{dX(t)}{dP(t,T)} \cdot dP(t,T) + \frac{1}{2}\frac{d^2X(t)}{dP(t,T)^2} \cdot dP(t,T)^2 =$$

$$= -\frac{1}{P(t,T)^2} \cdot dP(t,T) + \frac{1}{2}\left(-\frac{-2}{P(t,T)^3} \cdot \underbrace{\tilde{\sigma}(t)^2 P(t,T)^2 \left(dW^{Q_B}(t)\right)^2}_{dP(t,T)^2}\right) =$$

$$= -\frac{1}{P(t,T)^2}\cdot\left(r(t)P(t,T)dt + \tilde{\sigma}(t)P(t,T)dW^{Q_B}(t)\right) + \frac{\tilde{\sigma}(t)^2}{P(t,T)}dt =$$

$$= \frac{\tilde{\sigma}(t)^2 - r(t)}{P(t,T)}dt - \frac{\tilde{\sigma}(t)}{P(t,T)}\cdot dW^{Q_B}(t)$$

and substituting back into $d\left(\frac{B(t)}{P(t,T)}\right) = B(t)d\frac{1}{P(t,T)} + \frac{1}{P(t,T)}dB(t)$, and using that $dB(t) = B(t)r(t)dt$, we obtain

$$d\left(\frac{B(t)}{P(t,T)}\right) = B(t)\left(\frac{\tilde{\sigma}(t)^2 - r(t)}{P(t,T)}dt - \frac{\tilde{\sigma}(t)}{P(t,T)}\cdot dW^{Q_B}(t)\right) + \frac{r(t)B(t)}{P(t,T)}dt =$$

$$= \underbrace{\frac{\tilde{\sigma}(t)^2 B(t)}{P(t,T)}}_{\text{drift under } Q_B} dt - \underbrace{\frac{\tilde{\sigma}(t)B(t)}{P(t,T)}}_{\textit{diffusion}}\cdot dW^{Q_B}(t) \quad .$$

Now recall that $d\left(\frac{B(t)}{P(t,T)}\right)$ is a martingale under Q_T, hence under this measure (after a suitable change of measure from Q_B), this dynamics has no drift; therefore must be given by

$$d\left(\frac{B(t)}{P(t,T)}\right) = -\frac{\tilde{\sigma}(t)B(t)}{P(t,T)}\cdot dW^{Q_T}(t) \ ,$$

thus, $\frac{\tilde{\sigma}(t)^2 B(t)}{P(t,T)}dt - \frac{\tilde{\sigma}(t)B(t)}{P(t,T)}\cdot dW^{Q_B}(t) = -\frac{\tilde{\sigma}(t)B(t)}{P(t,T)}\cdot dW^{Q_T}(t)$, leading to

$$\boxed{dW^{Q_T}(t) = dW^{Q_B}(t) - \tilde{\sigma}(t)dt} \ .$$

Note that, remarkably, we have re-discovered the Sharpe Ration given by C-M-G Theorem:

$$\frac{\mu^{Q_B} - \mu^{Q_T}}{\sigma} = \frac{\frac{\tilde{\sigma}(t)^2 B(t)}{P(t,T)} - 0}{-\frac{\tilde{\sigma}(t)B(t)}{P(t,T)}} = -\tilde{\sigma}(t) \quad ,$$

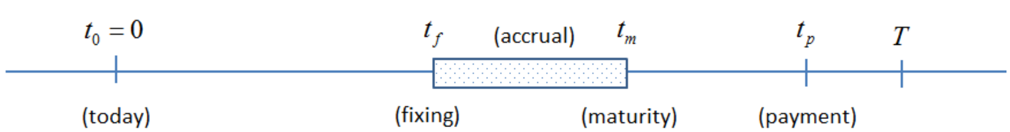
where we used the already found $\mu^{Q_B} = \frac{\tilde{\sigma}(t)^2 B(t)}{P(t,T)}$, $\mu^{Q_T} = 0$, and $\sigma = -\frac{\tilde{\sigma}(t)}{P(t,T)}$. In other words, the market price of risk is $-\tilde{\sigma}(t)$, where $\tilde{\sigma}$ is the instantaneous volatility of the ZC bond price $P(t,T)$. It can be shown that, in our Vasicek model, this volatility is given by $\tilde{\sigma}(t) = -H(t,T)\sigma$, with σ being the volatility of the short rate, and $H(t,T) = \frac{1-e^{-a(T-t)}}{a}$. Finally, we now can write the dynamics of the short rate under the T-forward measure Q_T as follows:

$$
\begin{aligned}
dr(t) &= a\left(b - r(t)\right)dt + \sigma dW^{Q_B}(t) = a\left(b - r(t)\right)dt + \sigma\left(dW^{Q_T}(t) - H(t,T)\sigma\right) = \\
&= a\left(b - r(t) - H(t,T)\sigma^2\right)dt + \sigma dW^{Q_T}(t)
\end{aligned}
$$

Note. Hypothetically, one can simulate (e.g., based on Euler' scheme) the dynamics of short rate, based on the above formula - that is, under the T-forward measure - and average (Monte Carlo) the outcomes at time T to compute $f(t,T) = E^{Q_T}\left[r(T) \mid \mathcal{F}_t\right]$. This would work; however, for every sought time T, one has to re-run the entire simulation, which is intimately related to Q_T (and T, implicitly, via the term $H(t,T)$). It resonates with the fact that we have one T-forward measure for every T; therefore we need a full simulation in each T-forward probability measure. This makes HJM or other methods more suitable for evolving the entire curve $f(t,T)$ indeed.

I. A Use-Case for T-Forward Measure: Cashflows

Suppose we have a simulator that produces future paths of some index short rate, and for each path, we have analytical means to compute the corresponding zero coupon bond prices. We want to evaluate a future cashflow (say, belonging to a swap) by Monte-Carlo. We set the stage by a timeline, as follows:



Today's time is $t_0 = 0$; t_f is the time of fixing the accrual rate; the accrual period is $\left[t_f, t_m\right]$; the cashflow payment occurs at t_p, and the swap maturity is at T. For each simulated path, we seek to evaluate the cashflow payment, given known discount bond prices $P_d(0,t)$ (extracted, say from an initial yield curve, and used for discounting the cashflows to the present) and given index bond prices

$P_i\left(t_1,t_2\right)$ (bond prices computed analytically on the path and corresponding to the simulated index short rate). Essentially, we have two rate dynamics, one for discounting (the funding rate) and the other for fixing (the index rate). Denote by $V\left(0\right)$ the present value of the cashflow occurring at time t_p. Under the T-forward measure, we can write the following derivation, which starts in a familiar way:

$$V\left(0\right)=P_d\left(0,T\right)E^{Q_T}\left[V\left[T\right]\ |\mathcal{F}_0\right]\underset{E^{Q_T}\left[V[T]\,|\mathcal{F}_{t_p}\right]=V[T]}{\overset{P_d(T,T)=1}{=}}P_d\left(0,T\right)E^{Q_T}\left[E^{Q_T}\left[\frac{V\left[T\right]}{P_d\left(T,T\right)}\ |\mathcal{F}_{t_p}\right]|\mathcal{F}_0\right]=$$

$$\underset{property}{\overset{martingale}{=}}P_d\left(0,T\right)E^{Q_T}\left[\frac{V\left[t_p\right]}{P_d\left(t_p,T\right)}\ |\mathcal{F}_0\right]\overset{V\left[t_p\right]=P_i\left(t_f,t_m\right)}{=}P_d\left(0,T\right)E^{Q_T}\left[\frac{P_i\left(t_f,t_m\right)}{P_d\left(t_p,T\right)}\ |\mathcal{F}_0\right]$$

Once we simulate/compute $\frac{P_i\left(t_f,t_m\right)}{P_d\left(t_p,T\right)}$, we can perform an average of these ratios across all the paths (in the Monte Carlo spirit) and discount the obtained average to the present using the discount bond $P_d\left(0,T\right)$. The result will be $V\left(0\right)$: today's value of the future cashflow.

II. Another Use-Case for T-Forward Measure: Swaplet

To exemplify the use of T -forward measure in derivative pricing in a simplest way, let's price a swaplet. Consider today's time $t_0=0$, the LIBOR rate L, and a payer swaplet that: (1) pays a chashflow at time S in the future, based on a forward interest rate $L\left(T,T,S\right)$ fixed at time T, against a notional N; and (2) receives a cashflow given by a fixed rate K, for the same period $\left[T,S\right]$, and against the same notional. We denote by τ the year fraction for period $\left[T,S\right]$, and by $V\left(t\right)$ the value of the contract at time t. Since the cashflow occurs at S (maturity), we will work under the S -forward measure.

It is clear that at the fixing date T, the value of the cashflow exchanged at time S is already known, equal to:

$$V\left(S\right)=N\cdot\tau\cdot\left(K-L\left(T,T,S\right)\right).$$

We have already established how to evaluate this quantity today, under the S -forward measure:

$$V\left(0\right)=P\left(0,S\right)\cdot E^{Q_S}\left[V\left(S\right)|\mathcal{F}_0\right]\ ,$$

where $P(0,S)$ is the price of the zero coupon bond paying LIBOR rate and maturing at S, and the expectation is taken under the S-forward measure. Here, we use rate L as both discount and index rate. Substituting $V(S)$ in the above we obtain:

$$\begin{aligned} V(0) &= P(0,S)\cdot E^{Q_S}\left[N\cdot\tau\cdot\left(K-L(T,T,S)\right)|\mathcal{F}_0\right]= \\ &= P(0,S)\cdot N\cdot\tau\cdot\left(K-E^{Q_S}\left[\left(L(T,T,S)\right)|\mathcal{F}_0\right]\right) \end{aligned}$$

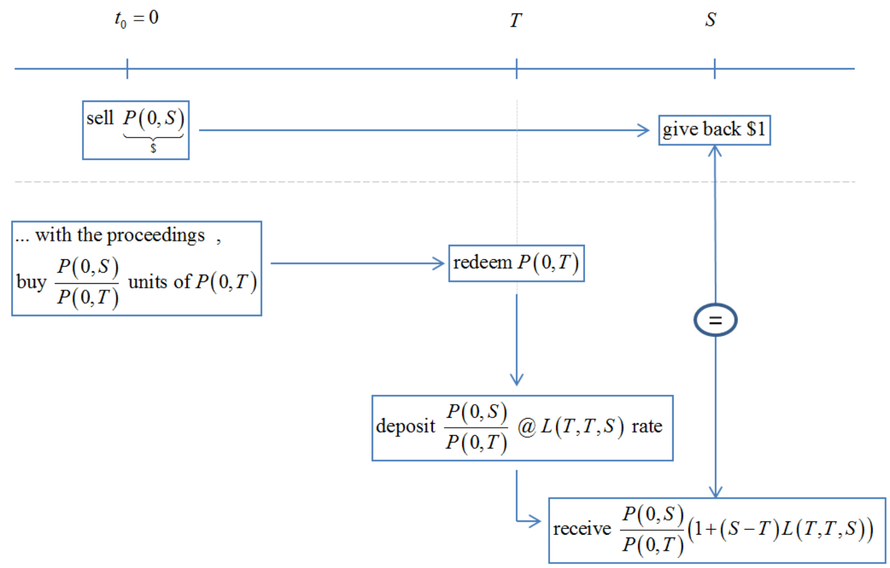
$L(\_,T,S)$ is a martingale under the S-forward measure, in that $E^{Q_S}\left[\left(L(T,T,S)\right)|\mathcal{F}_0\right]=L(0,T,S)$ which leads to the swaplet price today:

$$V(0)=P(0,S)\cdot N\cdot\tau\cdot\left(K-L(0,T,S)\right).$$

If we assume a simply compounded rate L, we have that $L(0,T,S)=\frac{1}{\tau}\left(\frac{P(0,T)}{P(0,S)}-1\right)$, where $P(0,\_)$ is a zero coupon bond that pays L, and which can be implied from the market. This can be justified by the following strategy – as in the figure below, we plan to:

1. Sell a ZC bond maturing at S, for the price of $P(0,S)$, and cash in the proceedings.
2. Buy units of ZC bond maturing at T: we can buy exactly $\frac{P(0,S)}{P(0,T)}$ units.
3. We let time pass, and at time T, we get back $\frac{P(0,S)}{P(0,T)}$ in cash, from the bond maturing at T.
4. We deposit $\frac{P(0,S)}{P(0,T)}$ cash for the period $[T,S]$ at LIBOR rate prevailing at time T: $L(T,T,S)$.
5. At time S we cash in the deposit with interest: $\frac{P(0,S)}{P(0,T)}\left(1+\tau\cdot L(T,T,S)\right)$.
6. The proceedings should precisely cover the payment of $1, due to the ZC bond bought at step 1.

Equating $1=\frac{P(0,S)}{P(0,T)}\left(1+\tau\cdot L(T,T,S)\right)$, then taking the expectation on both sides, and solving for L, leads to the expression $L(0,T,S)=\frac{1}{\tau}\left(\frac{P(0,T)}{P(0,S)}-1\right)$. The following is a schematics of the planned strategy:



$$E^T\left[\frac{P(0,S)}{P(0,T)}(1+(S-T)L(T,T,S)) \mid \mathcal{F}_0\right]=\frac{P(0,S)}{P(0,T)}(1+\tau\cdot L(0,T,S))=E^T[\$1 \mid \mathcal{F}_0]=1$$

Concluding, the swaplet price today is given by

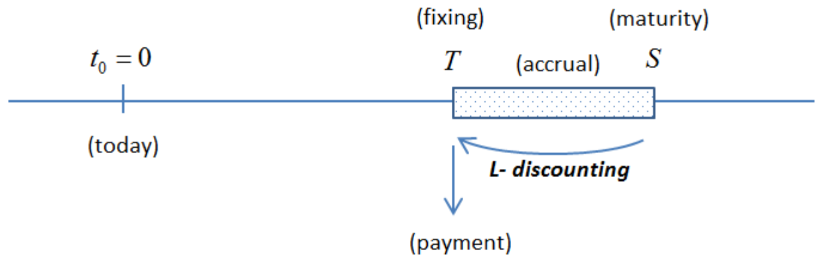
$$V(0)=P(0,S)\cdot N\cdot\left(\tau K-\frac{P(0,T)}{P(0,S)}+1\right).$$

We end this example with a note, for completeness. At a certain point in our presentation we stated that $L(\_,T,S)$ is a martingale under the S-forward measure, without justification. This is straightforward to observe. First note that there is nothing special about the initial time $t_0=0$. Indeed, the strategy that we used previously (with an implicit nonarbitrage argument) works for any time $t<T$. Therefore, we can generically state that $L(t,T,S)=\frac{1}{\tau}\left(\frac{P(t,T)}{P(t,S)}-1\right)$ for any $t<T$. Now, this is obvious a martingale under the S-forward measure, as $E^{Q_S}\left[\frac{P(t,T)}{P(t,S)} \mid \mathcal{F}_s\right]=\frac{P(s,T)}{P(s,S)}$ for any $s<t$, since $P(\_,T)$ is a tradeable asset expressed in units of bond numéraire. This shows that $E^{Q_S}\left[L(t,T,S) \mid \mathcal{F}_s\right]=L(s,T,S)$, that is, $L(t,T,S)$ is a martingale under the S-forward measure.

III. <u>Another Use-Case for T-Forward Measure: Vanilla FRA</u>

Another prototypical example is the pricing of a FRA, due to its peculiar payment convention. Consider the following timeline of a payer vanilla FRA, that pays the LIBOR index rate L for the accrual period

$[T, S]$ in the future, and receives a fixed reference rate K for the same period. The standard convention is to have the payment at the fixing time T, and this brings the necessity of discounting the payment amount for the year fraction $\tau = (T - S) / \text{days\_in\_year}$, at the same index rate.



For simplicity, we consider the FRA notional to be 1. Both the accrual and discounting assume a simply compounding index rate. Since the payment is done at time T, we start the valuation under the T-forward measure:

$$V(0) = P(0,T) E^{Q_T}\left[V(T) \mid \mathcal{F}_0\right] = P(0,T) E^{Q_T}\left[\frac{\tau \cdot (L(T,T,S) - K)}{1 + \tau L(T,T,S)} \mid \mathcal{F}_0\right]$$

Here, $P_L(0,T)$ is the discount bond maturing at T, and $L(T,T,S)$ is the LIBOR rate observed at fixing time T. Further, $\tau \cdot (L(T,T,S) - K)$ is the accrual amount and $(1 + \tau L(T,T,S))^{-1}$ is the discounting.

Obviously, the expectation is taken under the wrong forward measure. To fix this, we have to perform a change of measure and continue the derivation under the S-forward measure. More precisely, we seek a change of numéraire introduced in the previous section.

Denoting $Z(t) = E^{Q_T}\left[\frac{dQ_S}{dQ_T} \mid \mathcal{F}_t\right]$, we have $\frac{Z(t)}{Z(0)} = \frac{P(t,T)/P(t,S)}{P(0,T)/P(0,S)}$. Now recall the formula for the simply compounded LIBOR: $L(t,T,S) = \frac{1}{\tau}\left(\frac{P(t,T)}{P(t,S)} - 1\right)$, and further, $\frac{P(t,T)}{P(t,S)} = 1 + \tau L(t,T,S)$. This gives us the change of numéraire factor

$$\frac{Z(t)}{Z(0)} = \frac{1 + \tau L(t,T,S)}{1 + \tau L(0,T,S)}.$$

Back to our FRA pricing formula, we are now ready to perform the change of numéraire/measure from the T-forward measure to the S-forward measure:

$$V(0)=P(0,T)E^{Q_T}\left[\frac{\tau\cdot\left(L(T,T,S)-K\right)}{1+\tau L(T,T,S)}\,|\,\mathcal{F}_0\right]\overset{\text{change of numeraire}}{=}$$

$$=P(0,T)E^{Q_S}\left[\frac{\tau\cdot\left(L(T,T,S)-K\right)}{1+\tau L(T,T,S)}\cdot\frac{Z(T)}{Z(0)}\,|\,\mathcal{F}_0\right]=$$

$$=P(0,T)E^{Q_S}\left[\frac{\tau\cdot\left(L(T,T,S)-K\right)}{1+\tau L(T,T,S)}\cdot\frac{1+\tau L(T,T,S)}{1+\tau L(0,T,S)}\,|\,\mathcal{F}_0\right]=$$

$$=P(0,T)E^{Q_S}\left[\frac{\tau\cdot\left(L(T,T,S)-K\right)}{1+\tau L(0,T,S)}\,|\,\mathcal{F}_0\right]\overset{L\text{ is martingale under }Q_S}{=}$$

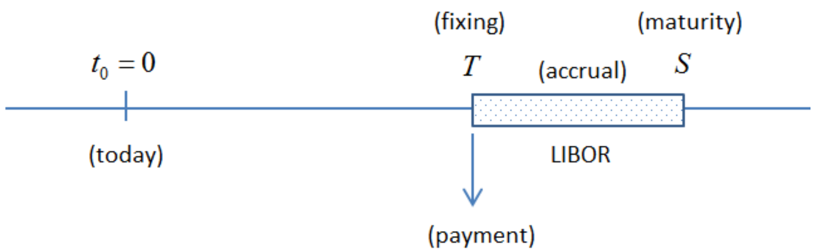
$$=P(0,T)\frac{\tau\cdot\left(L(0,T,S)-K\right)}{1+\tau L(0,T,S)}\overset{\frac{P(0,T)}{P(0,S)}=1+\tau L(0,T,S)}{=}P(0,S)\cdot\tau\cdot\left(L(0,T,S)-K\right)$$

which is the FRA price today, on a one-unit notional. For an arbitrary notional N, the price of this FRA becomes

$$V(0)=P(0,S)\cdot N\cdot\tau\cdot\left(L(0,T,S)-K\right)\ .$$

IV. Another Use-Case for Change of Numéraire: Convexity Adjustment for Libor in Arrears

We conclude this section with an application, that brings together the forward measures, the change of numéraire and Doléans-Dade exponential. Consider, yet again, a future cashflow payment as follows:



It resembles the cashflow of the floating leg of an FRA with the exception that the accrued amount is not discounted: is a *LIBOR in arrears* payment. The present value of this cashflow is given by

$$V(0)=P(0,T)E^{Q_T}\left[V(T)\,|\,\mathcal{F}_0\right]=P(0,T)E^{Q_T}\left[L(T,T,S)\,|\,\mathcal{F}_0\right]$$

Here $P(0,T)$ is the discount bond. Yet again, the expectation is taken under the wrong forward measure, reason for which we perform a change of numéraire, to express the expectation under the S-forward measure:

$$V(0)=P(0,T)E^{Q_S}\left[L(T,T,S)\frac{P(T,T)/P(T,S)}{P(0,T)/P(0,S)}\mid\mathcal{F}_0\right]=P(0,S)E^{Q_S}\left[\frac{L(T,T,S)}{P(T,S)}\mid\mathcal{F}_0\right].$$

Denoting $P_0(T,S)=\frac{P(0,S)}{P(0,T)}$ the current value of the forward discount factor, and equating the two expressions for $V(0)$ we obtain:

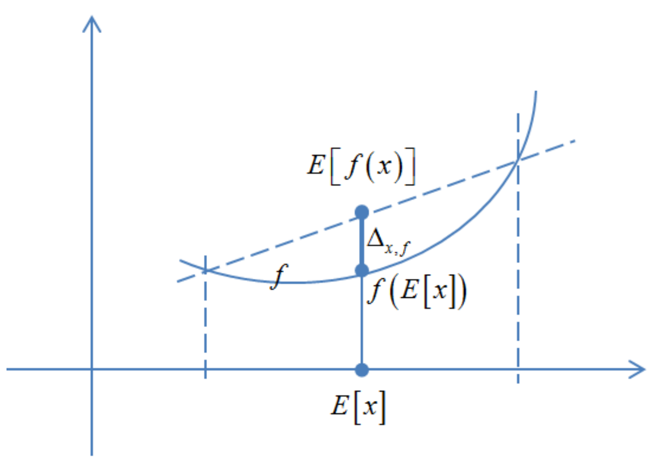
$$E^{Q_T}\left[L(T,T,S)\mid\mathcal{F}_0\right]=\frac{P(0,S)}{P(0,T)}E^{Q_S}\left[\frac{L(T,T,S)}{P(T,S)}\mid\mathcal{F}_0\right]=E^{Q_S}\left[L(T,T,S)\frac{P_0(T,S)}{P(T,S)}\mid\mathcal{F}_0\right]=$$

$$=E^{Q_S}\left[L(T,T,S)\mid\mathcal{F}_0\right]+\underbrace{E^{Q_S}\left[L(T,T,S)\left(\frac{P_0(T,S)}{P(T,S)}-1\right)\mid\mathcal{F}_0\right]}_{\Delta(T,S)}=$$

$$\overset{L(\_,T,S)\text{ is martingale under }Q_S}{=}L(0,T,S)+\Delta(T,S)$$

The term denoted by $\Delta(T,S)$ is called the convexity adjustment.

Note. The nomenclature "convexity adjustment" is inspired from Jensen's inequality, which in simple terms says that, for a convex functio f, we have $f\left(E\left[X\mid\mathcal{F}\right]\right)\le E\left[f(X)\mid\mathcal{F}\right]$, and Δ measures this deviation: $E\left[f(X)\mid\mathcal{F}\right]=f\left(E\left[X\mid\mathcal{F}\right]\right)+\Delta_{X,f}$. The relationship is intuitively depicted below.



Let's evaluate the convexity adjustment $\Delta(T,S)$, using the relation $L(t,T,S)=\frac{1}{\tau}\left(\frac{P(t,T)}{P(t,S)}-1\right)$:

$$\Delta(T,S)=E^{Q_S}\left[L(T,T,S)\frac{P_0(T,S)}{P(T,S)}\mid\mathcal{F}_0\right]-L(0,T,S)=$$

$$\overset{P(T,S)=\frac{1}{1+\tau L(T,T,S)}}{\underset{P_0(T,S)=\frac{1}{1+\tau L(0,T,S)}}{=}} E^{Q_S}\left[L(T,T,S)\frac{1+\tau L(T,T,S)}{1+\tau L(0,T,S)}\mid\mathcal{F}_0\right]-L(0,T,S)=$$

$$=\frac{1}{1+\tau L(0,T,S)}E^{Q_S}\left[L(T,T,S)+\tau L^2(T,T,S)\mid\mathcal{F}_0\right]-L(0,T,S)=$$

$$=\frac{E^{Q_S}\left[L(T,T,S)\mid\mathcal{F}_0\right]}{1+\tau L(0,T,S)}+\tau\frac{E^{Q_S}\left[L^2(T,T,S)\mid\mathcal{F}_0\right]}{1+\tau L(0,T,S)}-L(0,T,S)=$$

$$\overset{L\text{ ia martingale under }Q_S}{=}\frac{L(0,T,S)}{1+\tau L(0,T,S)}+\tau\frac{E^{Q_S}\left[L^2(T,T,S)\mid\mathcal{F}_0\right]}{1+\tau L(0,T,S)}-L(0,T,S)=$$

$$=\frac{\tau\left(E^{Q_S}\left[L^2(T,T,S)\mid\mathcal{F}_0\right]-L^2(0,T,S)\right)}{1+\tau L(0,T,S)}$$

Under the S-forward measure Q_S, $L(t,T,S)$ (L_t for brevity) is a martingale in its first argument, and under the assumption of a lognormal distribution, it accepts a martingale representation as $dL_t=\sigma_L L_t dW_t$, with W_t standard Brownian motion, and initial value L_0. We have already derived the solution of this SDE as a Doléans-Dade exponential: $L_t=L_0 e^{\sigma_L W_t-\frac{\sigma_L^2}{2}t}$. It remains to evaluate $E^{Q_S}\left[L_T^2\mid\mathcal{F}_0\right]$:

$$E^{Q_S}\left[L_T^2\mid\mathcal{F}_0\right]=E^{Q_S}\left[L_0^2 e^{2\sigma_L W_T-\sigma_L^2 T}\mid\mathcal{F}_0\right]=L_0^2\cdot e^{E\left[2\sigma_L W_T-\sigma_L^2 T\right]+\frac{1}{2}Var\left[2\sigma_L W_T-\sigma_L^2 T_0\right]}=L_0^2\cdot e^{-\sigma_L^2 T+\frac{1}{2}\cdot 4\sigma_L^2}=L_0^2\cdot e^{\sigma_L^2 T}$$

Finally, plugging in this expectation in the expression of $\Delta(T,S)$, we obtain

$$\Delta(T,S)=\frac{\tau\left(L^2(0,T,S)e^{\sigma_L^2 T}-L^2(0,T,S)\right)}{1+\tau L(0,T,S)}=\frac{\tau L(0,T,S)}{1+\tau L(0,T,S)}L(0,T,S)\left(e^{\sigma_L^2 T}-1\right).$$

Expanding the exponential to the first order, we can approximate the convexity adjustment by

$$\boxed{\Delta(T,S)\cong\frac{\tau L(0,T,S)}{1+\tau L(0,T,S)}L(0,T,S)\sigma_L^2 T}.$$

5 Epilogue

In this essay we shed light upon a fundamental technique in quantitative finance: change of measure. The main purpose was to build a clear intuition around the matter, with the drawback of missing the usual rigorousness and structure of a classical mathematical paper. Yet, we hope that the understanding and thoughts gained throughout this reading will allow practitioners to be more rigorous and insightful in their own writings. The following non-exhaustive list gives the main take-home ideas:

- ✓ Basic notions of Probability and Measure Theory are the building blocks for this technique. We have introduced the notions of sigma-algebras, probability spaces, random variables, and stochastic processes. Most importantly, filtrations and martingales are central concepts used across the entire essay.

- ✓ Brownian motions are "atomic" stochastic processes that fuel the dynamics of other processes that model market factors. A recurring subliminal message of this essay is that, most often, a change of measure can be viewed as a "change of Brownian motions." We don't have the means for controlling probabilities or probability spaces (in simulations); yet we can transform Brownian motions, and this is how we can change the measure.

- ✓ Girsanov Theorem, in its pure form, is a fairly abstract result; yet, with a few examples and discussions we hope to have built some intuition and provided an accessible meaning to this probability transformation. In particular, we found it useful to introduce so-called weight functions (Radon-Nikodym derivatives in disguise) as a tool for changing a probability measure to our liking: "tweaking" probabilities by adding more chance to some event occurring, and less to others, effectively allowed us to control the drift of a process. For example, we were able to turn a game of chance from an unfair setting to a fair game, without changing the payoff of the game itself.

- ✓ A simple example of change of measure is arguably the change of numéraire from a foreign currency to a domestic one – often used in pricing quanto instruments. We introduced this example before the formal definition of the risk-neutral measure (although this measure is mentioned in the example) because this application doesn't require a deep understanding of the more subtle notion of risk-neutrality.

- ✓ Finally, we introduced three most important probability measures: real, risk-neutral, and T-forward, and showed how one can change measure from real to risk-neutral, and from risk-neutral to T-forward. A main difference between these measures is that the real and risk-neutral mesures are universal (are defined on a given market) whereas the T-forward measure depends on a given instrument's time horizon: we have as many T-forward measures as times T's. We also reasoned about Monte Carlo simulations versus analytical derivations: the real and risk-neutral measures are used extensively in simulations, whereas the T-forward measure is more often used as a technique for simplifying analytical derivations in derivative pricing models.

We conclude with a justification for the title of this essay. In academic settings, instructors often make the curious analogy between change of measure and space travel: from Earth (our physical/real world) to some other planet, where habitants are insensitive to market risk. They even draw pictures of cosmonauts floating in the interstellar space between these worlds. And this is the only explanation that they give to the notion of risk-neutrality. In our opinion, there cannot be an analogy more absurd and wrong, and this analogy can only lead to a greater confusion and frustration for the student. The title of this essay, inspired by a most famous book of humor, is meant to ridicule and discourage such practice.

6 References

. . . any reference would defeat the purpose of this essay . . .

7 Copyright Information

This is a work of fiction. Names, symbols, formulas, results and outcomes either are the product of the author's imagination or are used fictitiously. Any resemblance to literature, published or otherwise, presentations, or ad hoc discussions is entirely coincidental.

No Greek letters were harmed in the writing of this essay.

For more information, address: nic.santean@gmail.com.

Innovation, Sciences et Développement économique Canada
Office de la propriété intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

Certificat d'enregistrement du **Droit d'auteur**

Certificate of Registration of **Copyright**

Ce certificat d'enregistrement est émis conformément aux articles 49 et 53 de la Loi sur le droit d'auteur. Le droit d'auteur sur l'oeuvre décrite ci-dessous, a été enregistré à la date d'enregistrement comme suit :

This Certificate of Registration is issued pursuant to sections 49 and 53 of the Copyright Act. The copyright in the work described below was registered on the date of registration as follows:

Date d'enregistrement - Date of Registration : **April 11, 2018**

Numéro d'enregistrement - Registration No. : **1148928**

Première publication - First Publication : **Unpublished**

Titre - Title : **The Hitchhiker's Guide to the Risk-Neutral Galaxy**

Catégorie - Category : **Literary**

Titulaire(s) - Owner(s) : **Nicolae Santean**
35 High Park Avenue, Apt. 1405
Toronto, Ontario
Canada, M6P 2R6

Auteur(s) - Author(s) : **Nicolae Santean**

Date d'émission du certificat - Date of Issuance of Certificate : **April 11, 2018**

Commissaire aux brevets Commissioner of Patents

(CIPO 00200)
24-05-16

Canada