The question

As I mentioned previously, Malliavin Calculus isn't a topic that can easily relate to everyday things. Most authors I've checked don't detail too much on why an object like it exists, or the Clark-Ocone formula. Instead, they go to the Hormander condition, which is meaningless for our purposes.

Malliavin Calculus is used in connection with stochastic processes and martingales, so we better explain those.

Stochastic processes

We can think of stochastic processes as a sequence of random variables that are indexed by time. I'll start with a very simple case, the one-dimensional random walk (also known as the drunkard's walk):

$$\mathbb{P}(\{X_i = +1\}) = 0.5$$
$$\mathbb{P}(\{X_i = -1\}) = 0.5$$
$$X_0 = 0$$
$$Z_n = \sum_{i=1}^{n} X_i$$

This is very simple: at time i, after Δt has passed, the drunkard take a step X_i, which can be either up or down with equal probability. The position Z at t is simply the result of all the steps the drunkard has taken before. As it stands, each movement "up" can be considered a "success", so this sum follows a binomial distribution centered at 0.

Let's consider something else: let the random walk take more frequent and shorter steps. In the language of the symbols above, Δt and X_i are made smaller. As the steps become infinitesimals, the process becomes continuous. This is how a path may look like when $\Delta t =$ 0.1, 0.01, 0.001:
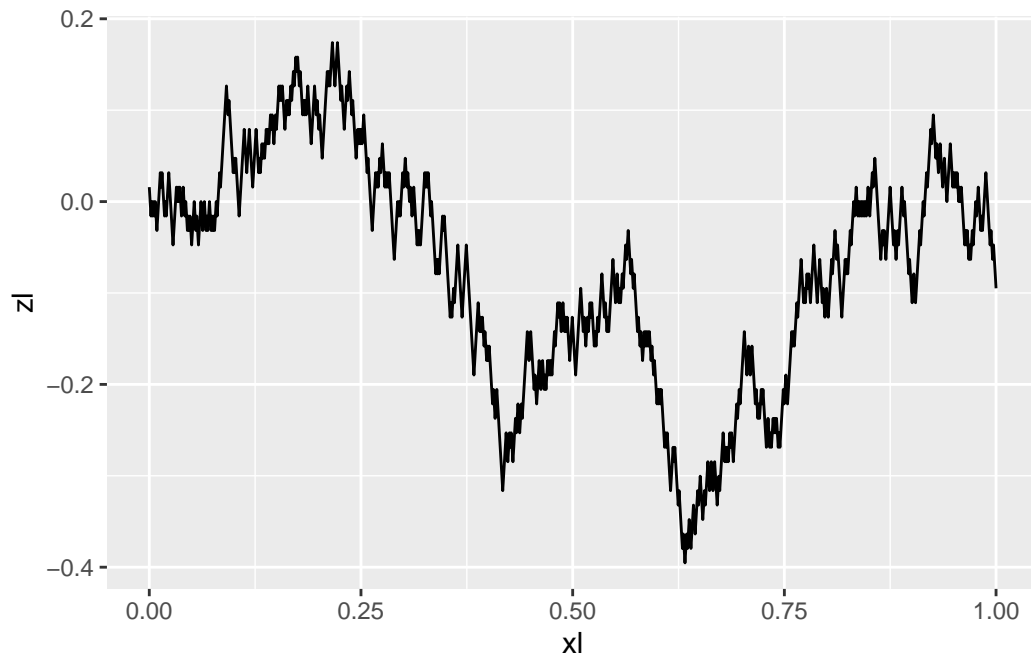
```
library(ggplot2)
library(rgl)
```

Warning: package 'rgl' was built under R version 4.3.2

```r
# Left
h <- 0.01
steps = 1 / h
xl <- seq(from = 0, to = 1, length.out=steps)
yl <- (rbinom(steps,1,0.5)*2-1)*sqrt(h)/2
zl <- cumsum(yl)
ggplot(mapping = aes(x = xl, y = zl)) + geom_line()
```
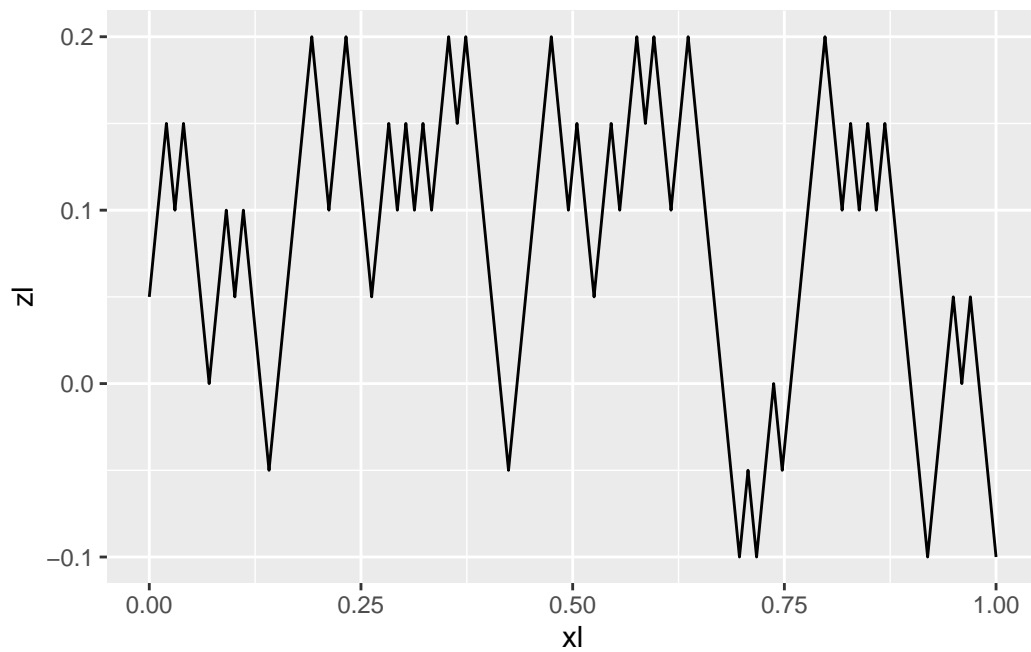


```r
# Middle
h <- 0.001
steps = 1 / h
xl <- seq(from = 0, to = 1, length.out=steps)
yl <- (rbinom(steps,1,0.5)*2-1)*sqrt(h)/2
zl <- cumsum(yl)
ggplot(mapping = aes(x = xl, y = zl)) + geom_line()
```

```
# Right
h <- 0.0001
steps = 1 / h
xl <- seq(from = 0, to = 1, length.out=steps)
yl <- (rbinom(steps,1,0.5)*2-1)*sqrt(h)/2
zl <- cumsum(yl)
ggplot(mapping = aes(x = xl, y = zl)) + geom_line()
```
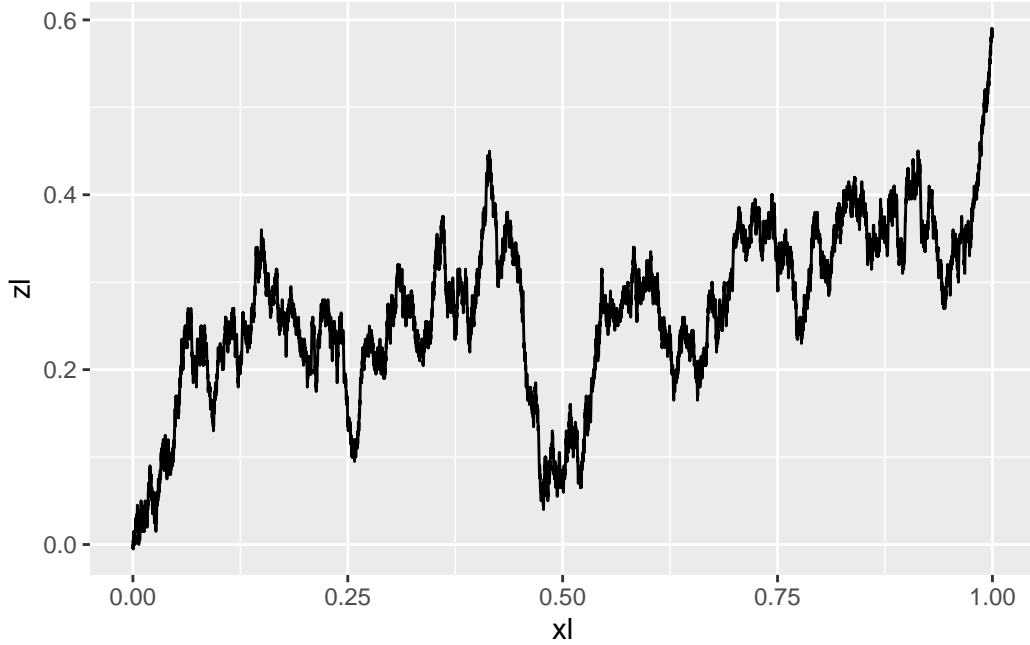
A one-dimensional random walk from 0 to t, in a similar fashion to the Central Limit Theorem, will converge to a normal variable:

$$Z_t \sim \mathcal{N}(\mu = 0, \sigma^2 = t)$$

These continuous-time stochastic processes are called Brownian motions or Wiener processes. They are denoted as B_t or W_t and fulfill a set of conditions:

- $B_0 = 0$

- Increments are independent, that is, $(B_{t+s} - B_t)$ doesn't depend on earlier values B_r, $r < t$

- Increments are Gaussian, that is, $(B_{t+s} - B_t) \sim \mathcal{N}(\mu = 0, \sigma^2 = s)$

- B_t is continuous at t. A bit circular, I know.

There's a catch, though. Even though they are continuous, these functions aren't differentiable. You kind of see it from the previous path graphs, with how jagged they become as the indices, or steps, shrink. It also makes sense if you think about the way we constructed the process: when you are at a point in time t, you don't know from where you randomly came. Did you arrive from a smaller or bigger value at $t - \Delta t$? There's no way to make sense of $\frac{\partial Z_t}{\partial t}$. This will become an issue later.

4

Ito Calculus

We have seen a very simple case of a stochastic process. What about more complicated random variables or functions of these variables?

We start from a simple substitution: we can't do $\frac{\partial Z_t}{\partial t}$ but we "can" pass the dt as a multiplication, like we do with total derivatives, and refer to just dZ_t. In general, we would like to express a minuscule change in the random variable being composed of something that changes very little in time and a little "noise". In short, something like this:

$$dZ_t = a(.)dt + b(.)dB_t$$

For starters, we can assume the functions a and b above are only functions of time. We then integrate over time and reach the definition of an **Ito process**:

$$\int_0^t dZ_t = Z_t - Z_0 = \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s$$

This representation is nice because the random variable Z_t has a mean that depends on something that changes along time, and has a variance tied to the noise. Having said that, we need a way to think about dB_t . For that, we can assume something simple, that instead of a continuum of dB_t, there's a list of discrete t_i and discrete increments ΔB_{t_i}, so now we deal with:

$$\sigma_{t_i} * (B_{t_i} - B_{t_{i-1}})$$

We can calculate the above because that difference of Brownian motions can be transformed into a normal distribution. To show how this help us, we can do the following calculation and confirm that the expected value depends only on the first term:

$$\mathbb{E}\left[\int_0^t \sigma_s \, dB_s\right] = \sum_{i=1}^n \sigma_{t_i} \mathbb{E}\left[B_{t_i} - B_{t_{i-1}}\right] = \sum_{i=1}^n \sigma_{t_i} \mathbb{E}\left[\mathcal{N}(0, t_i - t_{i-1})\right] = 0$$

Using this approach, we can show that Ito processes satisfy:

$$\mathbb{E}\left[X_t\right] = \int_0^t \mu_s \, ds$$

$$\mathbb{Var}\left[X_t\right] = \int_0^t \sigma_s^2 \, ds$$

Other ways we can deal with this B_t is with **Ito's isometry**, which allows you to exchange the dB_t into a more workable dt:

$$\mathbb{E}\left[\left(\int_0^t X_s \, dB_s\right)^2\right] = \mathbb{E}\left[\int_0^t X_s^2 \, ds\right]$$

With these tools, we can calculate a mean and a variance for every point in time in a process. But there's an even more powerful tool in our arsenal: **Ito's lemma**, or **Ito's formula**. In order to present this, we start with an Ito process for a random variable, say X_t:

$$dX_t = \mu_t \, dt + \sigma_t \, dB_t$$

$$X_t = X_0 + \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dB_s$$

Now, we want something more involved. Something like:

$$Y_t = f(t, X_t)$$

$$dY_t = a(t, X_t) \, dt + b(t, X_t) \, dB_t$$

with

$$dX_t = \mu_t \, dt + \sigma_t \, dB_s$$

Notice that we can't just do the typical chain rule from calculus. That is, we still can't do something like:

$$\frac{\partial Y_t}{\partial t} = \frac{\partial Y_t}{\partial X_t} \frac{\partial X_t}{\partial t}$$

But we can prove that there's an alternative chain rule we can apply, called **Ito's lemma**:

$$df(t, X_t) = \frac{\partial f}{\partial t} \, dt + \mu_t \frac{\partial f}{\partial x} \, dt + \sigma_t \frac{\partial f}{\partial x} \, dB_t + \frac{\sigma_t^2}{2} \frac{\partial^2 f}{\partial x^2} \, dt$$

$$df(t, X_t) = \underbrace{\left(\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{\sigma_t^2}{2} \frac{\partial^2 f}{\partial x^2}\right)}_{a(t, X_t)} dt + \underbrace{\left(\sigma_t \frac{\partial f}{\partial x}\right)}_{b(t, X_t)} dB_t$$

The first equation almost looks like a Taylor expansion up to the 2nd derivative, and in informal proofs that's what's going on. The second equation shows the explicit expressions for the $a(t, X_t)$ and $b(t, X_t)$ functions we mentioned before. This is great, because we can still split the "trend" component and the "noise" component for more complicated random processes. It also means that the "trend" is affected by how strong the "noise" is. Finally, notice that a non-random function with $\sigma_t = 0$ falls back to a classical total derivative and chain rule.

Martingales and final notions

Some stochastic processes present additional properties. One of those "some" are the martingales. They are either discrete or continuous-time processes that satisfy:

- $\mathbb{E}[X_t] < \infty, \forall t \geq 0$. This means that the process always has a value.
- $\mathbb{E}[X_{t+s}|X_t] = X_t, \forall t \leq s$. This means that the expected value for future realizations of the process is whatever value you have right now.

The drunkard's walk is an example of a martingale: at any time, it's equally likely it will go up or down, so the expected value is wherever you are at that point in time. Brownian motions are also martingales: increments follow a normal distribution, so the expected value of the increment is 0. Other examples of typical martingales are betting black or red at the roulette, or a stock price in the short term.

There's a result called the Martingale Representation Theorem, which states that any random variable X can be written in terms of another process C, with values known in advance, like so:

$$X = \mathbb{E}[X] + \int_0^\infty C_s \, dB_s$$

This is valuable for finance people because it means that any investment strategy (which is a random process) can be replicated with a different investment strategy (another process), but with lower volatility (and probably higher initial cost!). This theorem is less valuable because it doesn't give any means to calculate it, we can't apply a derivative and try to extract that dB_t

A second representation, from Rogers and Williams (2000), establishes that martingales can be represented as Ito integrals:

$$M_t = M_0 + \int_0^t \alpha_s \, dB_s$$

In a limited sense, α_t can be thought as $\frac{\partial M_t}{\partial t}$ because if we integrate it up to t we recover the final point, but it's not quite what we are looking for.

And it's always frustrating. We have been trying to find a way to calculate derivatives on these random processes, Brownian motions and martingales, but it's elusive. And the thing is, we kind of know that is possible. Indeed, let's go back to our first continuous random process, the one-dimensional Wiener process. We managed to reach the point were we can say that W_t had a normal distribution. So, nothing prevents us from taking a derivative over that, right?

$$W_t \sim \mathcal{N}(0, t)$$

$$W(x, t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}$$

$$\frac{\partial W}{\partial t} = \frac{e^{-x^2/(2t)} t^{-5/2} (x^2 - t)}{2\sqrt{2\pi}}$$

And there we have it. Ugly as it may be, we have constructed a derivative for how the density of these random variables evolve with time, although we don't know if it's even useful. Let's plot it.

```
options(rgl.useNULL=TRUE)
rgl::setupKnitr(autoprint = TRUE)
h = 200
Wt <- function(x, t) {
  if (t == 0) {
    return(0)
  } else {
    return(exp(-x^2/2/t)/sqrt(2*pi*t))
  }
}
dWt <- function(x, t) {
  if (t == 0) {
    return(0)
  } else {
    return(exp(-x^2/2/t)*t^(-5/2)*(x^2-t)/sqrt(8*pi))
  }
}
Wt <- Vectorize(Wt)
dWt <- Vectorize(dWt)

times <- seq(from=0, to=5, length.out=h)
xes <- seq(from=-5, to=5, length.out=h)

W <- outer(xes, times, Wt)
# persp(xes, times, W, col='white', shade=.1, theta = 150, phi = 15, ticktype='detailed',
```
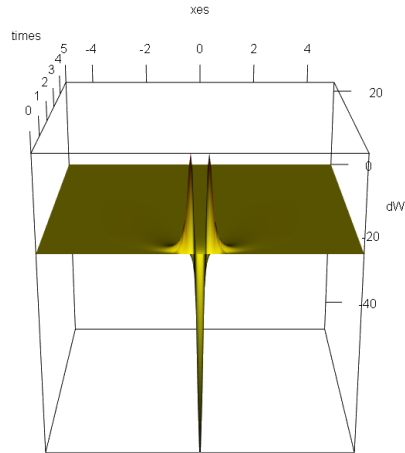
```
# open3d(windowRect=c(50,50,800,800))
palette <- colorRampPalette(c("blue", "green", "yellow", "red"))
col.table <- palette(256)
col.ind <- cut(W, 256)
persp3d(x=xes, y=times, z=W, col=col.table[col.ind])
```
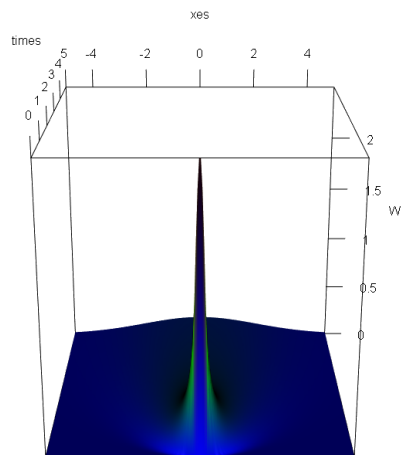


```
# rglwidget()

dW <- outer(xes, times, dWt)
# persp(xes, times, dW, col='white', shade=.1, theta = 150, phi = 15, ticktype='detailed',
col.ind <- cut(dW, 256)
persp3d(x=xes, y=times, z=dW, col=col.table[col.ind])
```

```
# rglwidget()
```

Conclusion

At this point, we have seen enough. We know what we want: a way to calculate derivatives over random variables and Wiener processes "in a certain sense" that's useful. That's what Malliavin calculus is set to do.

Rogers, L. C. G., and David Williams. 2000. *Diffusions, Markov Processes and Martingales.* 2nd ed. Vol. 2. Cambridge Mathematical Library. Cambridge University Press. https://doi.org/10.1017/CBO9780511805141.