Finance in the frequency domain

Medlemsmøte i Den Norske Aktuarforening

Berent Å. S. Lunde

University of Stavanger
Department of Mathematics and Physics

3. May 2018

Agenda

The frequency domain

Why?

- For most people: a place to decode a signal.
- For us: a fantastic place to combine randomness!
- For some of us: a place to prove properties about randomness.

What is it?

- The domain of the Fourier transform.

Take away:

- Properties to exploit in daily life.
- Basic understanding of the Fourier transform.
- An idea of what can be done.

The characteristic function and the frequency domain

Frequency: the number of occurrences of a repeating event per unit of time.
The characteristic function takes a frequency, s, as argument:

$$\varphi_X(s) = \mathbb{E}\left[e^{isx}\right] = \int e^{isx} p_X(x)\, dx$$

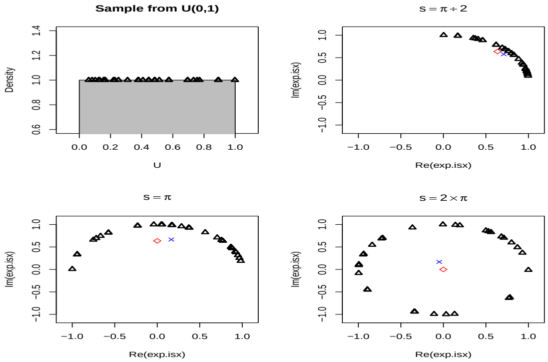
- s is the *angular frequency* $s = 2\pi\tau$, τ is the number of rotations per unit of time.
- The average of X wrapped around the unit circle with *angular frequency* s.
- $\varphi_X(s)$ is called the *Fourier transform* of $p_X(x)$.

The characteristic function and the frequency domain

The characteristic function takes a frequency, s, as argument:

$$\varphi_X(s) = \mathbb{E}\left[e^{isx}\right] = \int e^{isx} p_X(x)\, dx$$

- The average of X wrapped around the unit circle with *angular frequency* s.



Properties of the characteristic function

General property, if X_1 and X_2 are independent, then

$$\varphi_{(X_1+X_2)}(s) = \varphi_{X_1}(s) \times \varphi_{X_2}(s).$$

We will exploit this extensively!
We can obtain the density and distribution functions:

$$p_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(s) e^{-isx} ds,$$

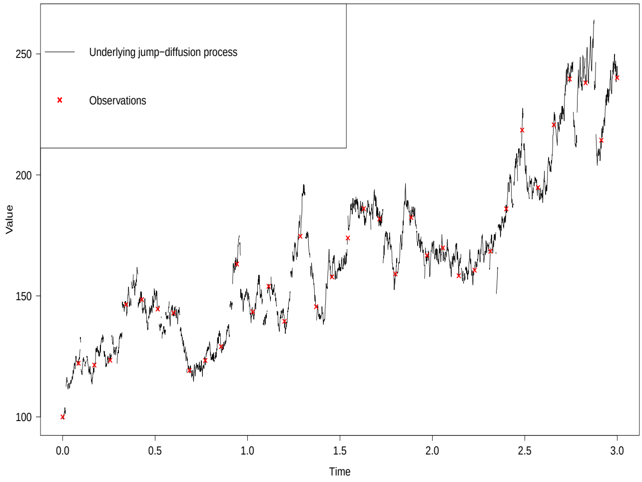
$$\mathbb{P}(X \leq x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{\mathbb{I}\mathrm{m}\left[e^{itx}\varphi_X(s)\right]}{s} ds.$$

The characteristic function for the sum of random variables

$$
\begin{tikzcd}
p_{X_1}(x) \arrow[r, "\text{FT}"] & \varphi_{X_1}(s) \arrow[ddr, "\text{Product}"] & & \\
p_{X_2}(x) \arrow[r] & \varphi_{X_2}(s) \arrow[dr] & & \\
\vdots & \vdots & \varphi_{\sum X_i}(s) \arrow[r, "\text{IFT}"] & p_{\sum X_i}(x) \\
p_{X_n}(x) \arrow[r] & \varphi_{X_n}(s) \arrow[ur] & &
\end{tikzcd}
$$

- Fourier transform (FT) integrals can typically be done analytically.
- Inverse Fourier transform (IFT) must be done numerically.
- The alternative is an $n-1$ dimensional integral, to be done numerically.

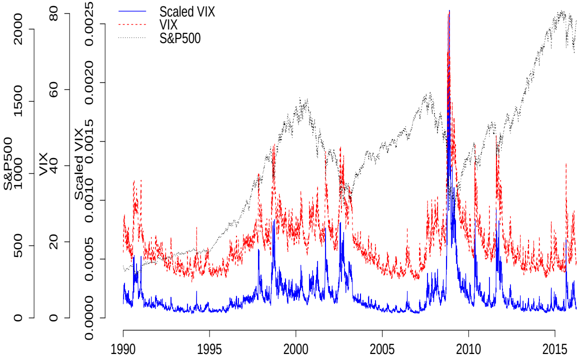
Example 1: Jump diffusions



Financial crashes

Bubbles are defined and explained after they occur.

Q: Is it possible, from the dynamics of stock prices, to correctly classify a bubble?



The empirical mystery: dynamics of financial bubbles

An intriguing idea:
Does financial bubbles grow in such a way that a crash is certain (at some point in the finite future)?

- Rephrase: does financial bubbles grow "super-exponentially"?
- Rephrased mathematically: for the deterministic part of dS_t:

$$\text{deterministic}[dS_t] = rS_t^{\alpha}dt,$$

  is $\alpha > 1$? (Compare this to exponential growth: $dS = rSdt$)

Goal: given time-series observations of a bubble - estimate α.

Jump diffusions

We must control for the growth that is due to randomness. Relatively accepted model: Merton jump diffusion A stochastic differential equation (SDE) with jumps

$$\frac{dS_t}{S_{t-}} = (r - \lambda k)dt + \sigma dW_t + (Y_t - 1)dN_t,$$

- Several sources of randomness: W_t, Y_t, N_t.
- Extend this with α! (can be done in several different ways)

SDE's with jumps

- SDE's are difficult, no general way to solve them. Adding jumps to the problem only increases the complexity.

Frequency domain to the rescue!

- Realize that we are dealing with several sources of independent randomness: W_t, Y_t, N_t.

$$
\begin{tikzcd}
dS_t \arrow[r] & S_T \approx \tilde{S}_T \arrow[r] & \varphi_{\tilde{S}_T}(s) \arrow[dr] & & \\
Y_t \arrow[dr] & & & \varphi_{(\tilde{S}_T+Z_T)}(s) \arrow[r] & p_{(\tilde{S}_T+Z_T)}(s) \\
& Z_T = \sum_{i=1}^{N_T} Y_i \arrow[r] & \varphi_{Z_t}(s) \arrow[ur] & & \\
N_t \arrow[ur] & & & &
\end{tikzcd}
$$

The inverse Fourier transform

One more difficulty - the IFT:

$$p_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(s) e^{-isx} ds,$$

must be evaluated numerically.

- Integrals approximated as sums.
- Exact arithmetic versus floating point.
- Problems for low-density regions.

Saddlepoint adjusted IFT: $K(s) = \log \varphi(-is)$

$$\frac{\exp\left[K_X(\hat{s}) - \hat{s}x\right]}{2\pi\sqrt{\left|K_X''(\hat{s})\right|}} \int \varphi_{\tilde{X}(\hat{\tau})}(s) ds.$$

Empirical results:

Model		Parameters							Statistics		
		r	σ	α	β	λ	μ	ν	$l(\hat{\theta}; \mathbf{x})$	D	p-value
					SSE bubble of 07						
GBM	est	0.6268	0.2629						1796.7		0.0013
	se	0.1604	0.0071								
CEV	est	0.4718	0.0118	1.4072					1826.2	59	0.0011
	se	0.1478	0.0021	0.0244							
nlModel 1	est	0.0249	0.0112	1.4132					1827.6	61.8	0.0030
	se	0.0082	0.0019	0.0235							
nlModel 2	est	0.0001	0.0120	2.0744	1.4046				1828.9	64.4	0.0020
	se	0.0005	0.0022	0.3920	0.0247						
MJD	est	0.6261	0.1694			92.1	-0.0039	0.0202	1840.9	88.4	0.7645
	se	0.1584	0.0272			64.9	0.0027	0.0051			
CEVJD	est	0.4356	0.0128	1.3769		13.6	-0.0094	0.0345	1851.8	110.2	0.0179
	se	0.1525	0.0033	0.0345		8.2	0.0086	0.0085			

Recap from example 1

- The frequency domain provided a solution.
- Created a fast, exact, and robust inversion algorithm which is generic.
- Difficult to ”diagnose” a bubble from the dynamics alone.

Example 2: Aggregating pricing models

The classical approach to pricing

- Given historical claims-data, a portfolio of customers, and a target/KPI to attain: find the price for each customer.
- Classical approach:
  1. Frequency times severity model.
  2. Parametrise the model to the historical data.
  3. Scale the pure premiums (given by the model on the customers,) to expect to reach the KPI target.

The classical (mathematical) approach to pricing

Find the price for each customer:

- Given historical claims-data, Y,
- a portfolio of customers $\mathbb{V} = \sum_{i=1}^{n} V_i$, with risk factors x_i
- and a KPI: $g(\mathbb{V}) = k$, for example loss ratio: $\frac{\mathbb{V}}{\Pi} = 0.5$, combined ratio, or return on equity.

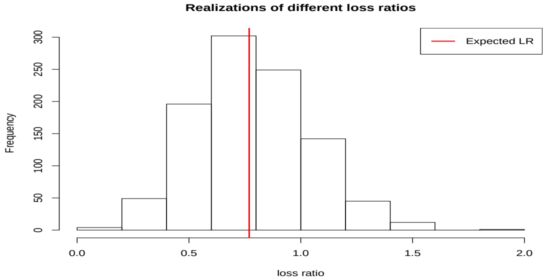
The classical approach:

1. $V(x) = \sum_{j=0}^{N(x)} Y_j(x)$ - leads to a frequency $\times$ severity approach.
2. One model for $N(x)$ (typically Poisson), and one for $Y(x)$ (typically Gamma). Then $\mathbb{E}[V(x)] = \mathbb{E}[N(x)] \times \mathbb{E}[Y(x)]$.
3. Parametrise the models to the historical data.
4. Solve $\mathbb{E}\left[g(\mathbb{V})\right] = k$ for some scaling parameter α, for example, for the loss ratio:

$$\frac{\sum_{i=1}^{n} \mathbb{E}[N; x_i]\mathbb{E}[Y; x_i]}{\sum_{i=1}^{n} P(\alpha, x_i)} = k$$

$(P(\alpha, x_i)$ depends on how the premium is structured).

Attaining the KPI



- An α is chosen to attain $g(\mathbb{V}) = k$.
- Important to remember that $g(\mathbb{V})$ is stochastic.
- What if the realization of $g(\mathbb{V}) > k$? Probably, someone will complain.

Attaining the KPI

If g is monotone in $\mathbb{V}$ and continuous, then

- Realize that all the randomness in $g(\mathbb{V})$ stems from $\mathbb{V}$.
- $\mathbb{V}$ is a sum of independent risks V_i.
- The randomness of V_i is already modelled as $V(x) = \sum_{j=0}^{N(x_i)} Y_j(x)$.
- This begs for a treatment in the frequency domain:

A sum of random sums

- Utilizing the frequency domain

$$
\begin{tikzcd}
Y(x_1) \arrow[dr] & & & \\
& \varphi_{V(x_1)}(s) \arrow[ddr] & & \\
N(x_1) \arrow[ur] & & & \\
\vdots & \vdots & \varphi_{\mathbb{V}}(s) \arrow[r] & p_{\mathbb{V}}(s) \\
Y(x_n) \arrow[dr] & & & \\
& \varphi_{V(x_n)}(s) \arrow[uur] & & \\
N(x_n) \arrow[ur] & & &
\end{tikzcd}
$$

- The density of the KPI can be retrieved by:

$$p_{g(\mathbb{V})}(k) = p_{\mathbb{V}}\left(g^{-1}(k)\right)\left|\frac{d}{dk}g^{-1}(k)\right|$$

Premium principles

We can then define premium principles w.r.t. the portfolio in a two step procedure:

1. Find sum of premiums Π s.t. some equation with a functional on the space of the KPI is satisfied. For example

   $$\mathbb{E}\left[g\left(\mathbb{V},\Pi\right)\right]=k, \text{ (This is what you do every day.)}$$

   or cooler (if g monotone increasing in $\mathbb{V}$):

   $$\inf\left\{\Pi:\mathbb{P}\left(g\left(\mathbb{V},\Pi\right)\leq k\right)<1-\alpha\right\}, \text{ for chosen } \alpha.$$

   For example: smallest premium such that 95% sure that loss-ratio will be less than 0.5.
2. Distribute the premium to the risks according to some structure. Most typically: $H(x_i)=\frac{\mathbb{E}[V(x_i)]}{\mathbb{E}[\mathbb{V}]}\Pi$.

Recap from example 2

- The frequency domain is an ideal tool for a portfolio.
- Retrieval of the KPI density.
- Extend the definition of premium principles to consider the space of the KPI. Leads to more control over final product.
- The best part: this should not lead to more work for the actuary!

Example 3: Model selection

Preliminary: Linear vs. polynomial regression

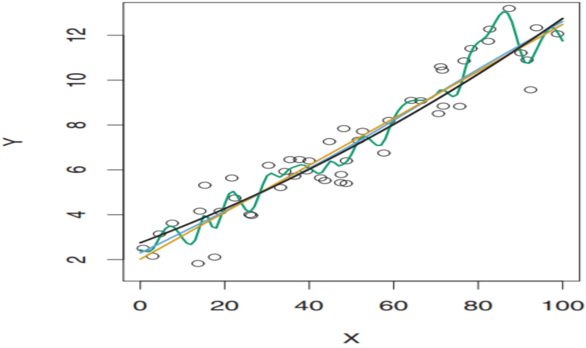


Figure: Figure from ”Introduction to statistical learning” p.33

- Which is better?
- Motivates a ”Model selection criteria”.
- Especially relevant for machine-learning methods where the functional form changes during training.

Model selection criteria

$$T = \mathbb{E}_x \left[\mathbb{E}_y \left[\mathcal{L} \left(y, f(\cdot; \hat{\theta}(x)) \right) \right] \right]$$

The training loss is a *biased* estimator of T. Many information criteria tries to correct for this bias

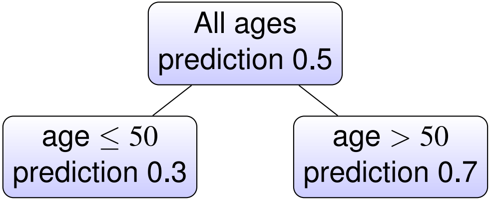
- Asymptotic criteria: AIC, AICc, BIC, DIC, TIC, NIC, ...
- Data criteria: Cross Validation, Validation dataset, EIC

Also possible to approximate T using the resampling distribution:

$$T \approx \int \int \mathcal{L} \left(y, f(\cdot; \hat{\theta}(x)) \right) d\hat{F}(x) d\hat{F}(y)$$
$$\approx \frac{1}{B} \sum_{b=1}^{B} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L} \left(y_i, f(\cdot; \hat{\theta}(y_i^b)) \right)$$

Model selection: tree splitting

When training a tree, the functional form constantly changes: split or no split?



Which model selection criteria?

- Asymptotic criteria: assumptions may not hold.
- Data criteria: computationally expensive - Cross validation is the norm.

Gradient tree loss

For gradient tree boosting, the relevant part of the approximate loss in a node is:

$$\mathcal{L} = \sum_{i=1}^{n} g_i^{(1)} w(y^{(2)}) + \frac{1}{2} h_i^{(1)} w(y^{(2)})^2$$

where $w(y^{(2)})$ is the prediction in the node, trained from an independent (re)sample $y^{(2)} = \left\{ g_i^{(2)}, h_i^{(2)} \right\}_{i=1}^{n}$:

$$g_i = \frac{\partial}{\partial \hat{y}} \mathcal{L}(y_i, \hat{y}), \qquad h_i = \frac{\partial^2}{\partial \hat{y}^2} \mathcal{L}(y_i, \hat{y})$$

$$w(y^{(2)}) = -\frac{\sum_{i=1}^{n} g_i^{(2)}}{\sum_{i=1}^{n} h_i^{(2)}} = -\frac{G^{(2)}}{H^{(2)}}.$$

- We have sums of independent random variables.

Resampling in the frequency domain

The characteristic function of drawing one observation from the original sample is

$$\varphi_{g^b,h^b}(s,v) = \frac{1}{n}\sum_{j=1}^{n} e^{isg_j+ivh_j}.$$

We draw n such observations with replacement (product rule):

$$\varphi_{G^b,H^b}(s,v) = \prod_{j=1}^{n} \varphi_{g^b,h^b}(s,v) = \left[\varphi_{g^b,h^b}(s,v)\right]^n$$

An integrated loss criterion

Strategy: instead of bootstrapping:

1. View the resampling in the frequency domain.
2. Combine terms.
3. Apply LOTUS with the empirical density of G and H (found after inversion):

$$T \approx \int \int \mathcal{L}\left(y, f(\cdot;\hat{\theta}(x))\right) d\hat{F}(x) d\hat{F}(y)$$
$$\approx \int \left[\bar{g}\frac{G}{H} + \frac{1}{2}\bar{h}\left(\frac{G}{H}\right)^2\right] d\hat{F}\left(G, H\right)$$

here $\bar{g}$ and $\bar{h}$ are the means from the original sample, and $d\hat{F}\left(G, H\right)$ is approximated as $\mathtt{spa}\left(\hat{f}, (G, H)\right) dGdH$.

Example

- Loss function: log loss.

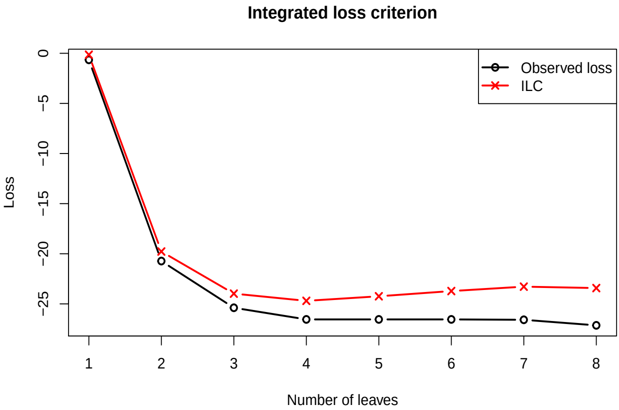


Figure: The ILC correctly chooses the model with 4 leaves.

Recap from example 3

- Model selection criteria to find the model that generalizes the best.
- The functional form of a tree changes constantly during training - repeated model selection takes place.
- Asymptotic information criteria relies heavily on assumptions.
- Data driven approaches such as cross-validation are computationally costly.
- Find a middle-ground with a data driven approach that is computationally efficient due to resampling in the frequency domain - also, should be more efficient than cross validation, as it utilizes all information in the data.
- Makes gradient tree boosting fairly automatic.

Finance in the frequency domain: summary

- The frequency domain is the domain of the Fourier transform.
- Fourier transform is the expectation of the random variable wrapped around the unit circle with a given frequency.
- Independent randomness combines very well in the frequency domain.
- It is ideal to combine randomness in the frequency domain to avoid high-dimensional numerical integration.
- It allowed us to combine the jumps and the continuous randomness in the SDE (and actually also the Itô-Taylor expansion to the solution of the pure SDE).
- It allowed aggregation of risks in a portfolio, which allows for extending regular premium-principles to have control over KPI targets.
- It allowed for fast resampling by considering drawing from the data with replacement in the frequency domain.

End

Next time when you have a problem with multiple sources of randomness - perhaps it can be (partly) solved by considering the frequency domain?

Thank you for your attention!