Extreme Values in Financial Statistics

Master Thesis Project done at EPFL
towards the French 'Diplôme d'Ingénieur' degree
under the exchange agreement EPFL-ENSEEIHT

Thèse de Master
présentée le 15 Août 2015
sous la supervision de

Prof. Stephan Morgenthaler,
Chaire de Statistique Appliquée
École Polytechnique Fédérale de Lausanne

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

pour l'obtention du Diplôme d'Ingénieur ENSEEIHT
en Mathématiques Appliquées et informatique
par

Killian Martin–Horgassan

Lausanne, EPFL, 2015

Mihi cura futuri
— Ovide, Métamorphoses, 13, 363

To my friends and loved ones…

Acknowledgements

TO BE FILLED

Lausanne, 14 Août 2015 K. M–H.

Preface

A preface is not mandatory. It would typically be written by some other person (eg your thesis director).
TO BE FILLED

Lausanne, 14 Août 2015 K. M–H.

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Key words:

Résumé

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Mots clefs :

Contents

List of Figures

List of Tables

1 Introduction

1.1 A few words to set the scene

In real life, it is not uncommon to have at one's disposal data about a phenomenon occurring through time. It may be as simple as daily rainfall data in a cit for the past two years, or it could be the weekly opening prices of a stock for the past decade.

Most of the time, people would like to use the data at their disposal to make predictions to answer questions, from the prosaic ones such as 'Will it rain tomorrow ?' to more consequential ones such as 'Will I make a profit if I cling to my shares today and sell them only tomorrow ?'. Of course, those are only vaguely worded questions : it is impossible to answer them satisfactorily without knowing the context, the objectives etc. behind them.

Yet, what these questions have in common is that they focus on the normal 'behaviour' that is to be expected in the future. Depending on the specific issue that is considered, the 'average behaviour' may not be the most interesting thing. For instance, suppose that a government wants to build a network of dams[1]. The dams are meant to protect the country from future floods for the next one hundred years, therefore the question that needs to be answered is one of 'worst case event' : "Over the next century, how severe may be the worst flood ?".

Extreme events are the kind of events we will be interested in this master thesis project. Although Extreme Value Theory has applications in many fields[2], we will here apply it more specifically to financial data.

1.2 Formalising the settings

Let $(X_n)_{n\geq 0}$ be a sequence of independent identically distributed random variables with common cumulative distribution function F_X. The sequence of maxima is defined by M_0=X_0 and $\forall n \geq 1$, M_n=$\max_{0\leq i\leq n}(X_i)$. We would like to determine the limiting distribution of the

[1]As was done in The Netherlands beginning in the fifties
[2]including climate science, seismology, insurance etc.

sequence $(M_n)_{n\geq 0}$.[3] This is a matter that will keep us busy quite a long time but the first thing to do is to re-formulate it.

Indeed, let us do a quick and simple computation :

$$\begin{aligned} F_n(t) &= \Pr(\{M_n \leq t\}) \\ &= \Pr(\{\max_{0\leq i\leq n}(X_i) \leq t\}) \\ &= \Pr(\{X_1 \leq t\} \cap \cdots \cap \{X_n \leq t\}) \\ &= (F_X(t))^n \end{aligned} \tag{1.1}$$

Here we see that not much information will be drawn from this result by taking the limit n→ +∞. The limiting distribution will be degenerate. Indeed, let us consider the upper end-point of F_X[4], z^+. Then,

$$\begin{aligned} &\forall z < z^+ \lim_{z\to\infty} F_n(z) = 0 \\ &\forall z \geq z^+ \lim_{z\to\infty} F_n(z) = 1 \end{aligned} \tag{1.2}$$

It turns out we cannot use the limiting distribution directly. A common approach[5] is to consider a sequence of the maxima, centred and normalised.

We will thus consider in all what follows the sequence defined by $(M_n^*)_{n\geq 0} = (\frac{M_n - b_n}{a_n})_{n\geq 0}$ where $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ are a sequence of real numbers and positive real numbers respectively. Finding a result on whether such a sequence admits a limiting distributions, and the conditions under which the result holds, will be one of our goals.

[3] If we can determine the limiting distribution of the maxima from the data, then we will have a means to make predictions on the occurrence of future extreme events.

[4] that is the smallest z such that $F_X(z)$ be equal to one. For the Normal distribution, z will be $+\infty$, by contrast for a continuous Uniform Distribution $U([a,b])$ it will be b.

[5] adopted by the mathematicians that laid the grounds for EVT

2 Investigating results on the limiting distribution

2.1 Playing with the (original) sequence of maxima

Here, we will generate finite size sequences (N = 10000) of independent identically distributed random variables following respectively :

- a standard Normal Distribution $\mathcal{N}(0,1)$
- a Cauchy Distribution *Cauchy(0,1)*
- an Exponential Distribution *Exp(1)*

We will compute the sequence of maxima, neither centred nor normalised, and draw the scatter plot as well as the plot of the maxima M_n as a function of the time steps n. We will also draw the $\frac{1}{n}$-quantiles of the distributions (distributions of the sample, not of the maxima) as a function of the time steps n. This will lead us to make an interesting observation.

2.1.1 Sample following a Normal distribution

Computing the quantiles The Normal distribution is a particular case because, unlike in the cases of the Cauchy and the Exponential distribution, there is no explicit form to the cumulative distribution function. We will thus use a "well-known"[1] inequality, holding $\forall\, t > 0$:

$$(\frac{1}{t} - \frac{1}{t^3}) * \frac{\exp(-\frac{t^2}{2})}{\sqrt{2*\pi}} < 1 - \Phi(t) < \frac{1}{t} * \frac{\exp(-\frac{t^2}{2})}{\sqrt{2*\pi}} \tag{2.1}$$

[1] Many textbooks mention it, though it is not necessarily what springs to the mind when thinking about the properties of Gaussian RVs.

From there, it is easy to see that the following holds :

$$1-\Phi(t) \sim_{t\to+\infty} \frac{1}{t} * \frac{\exp(-\frac{t^2}{2})}{\sqrt{2*\pi}} \tag{2.2}$$

When n grows large, the $\frac{1}{n}$-quantile grows very large so it is valid to replace $1-\Phi(t)$ by its equivalent in the equation satisfied by the quantiles :

$$\begin{aligned} & F_X(q_{\frac{1}{n}}) = 1 - \frac{1}{n} \\ \iff & \frac{1}{q_{\frac{1}{n}}} * \frac{\exp(-\frac{q_{\frac{1}{n}}^2}{2})}{\sqrt{2*\pi}} = \frac{1}{n} \\ \iff & \log(q_{\frac{1}{n}}) + \log(\exp(-\frac{q_{\frac{1}{n}}^2}{2})) + \log(\sqrt{2*\pi}) = \log(n) \end{aligned} \tag{2.3}$$

This equation cannot be solved analytically, we will resolve it iteratively. Th starting point is $\log(n) = \frac{t_0^2}{2}$, which gives us $t_0 = \sqrt{(2*\log(n))}$. If we then run the Newton-Raphson algorithm, we see that the corrections to t_0 from the next iterations are small enough that we can keep t_0 as solution.[2].

Figure 2.1 – Below, a realisation of the sequence of maxima for i.i.d. standard unit Gaussian RVs

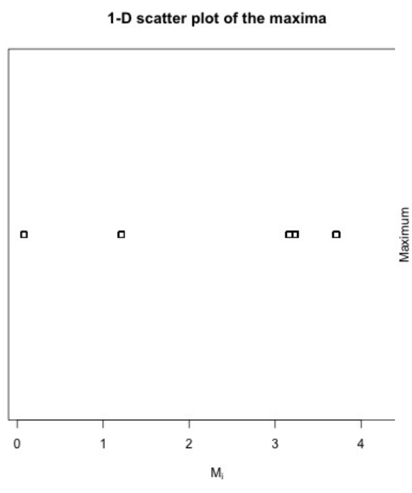


Figure 2.2 – Scatter Plot of the Maxima, n = 10000

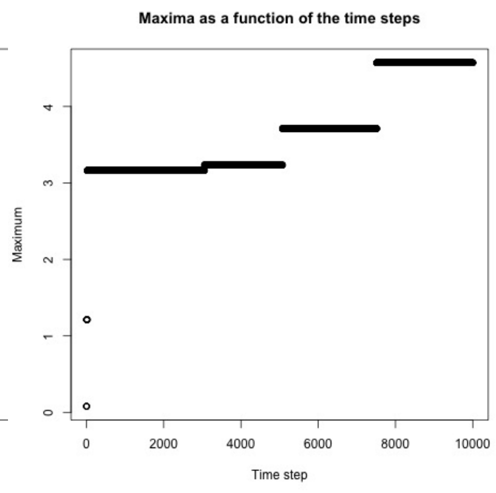


Figure 2.3 – Maxima against the time steps

[2]See the fourth of the figures below

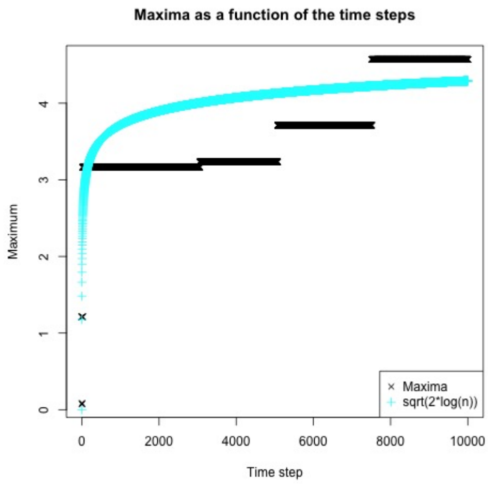


Figure 2.4 – Maxima against the time steps and function n→ $\sqrt{2 * \log(n)}$

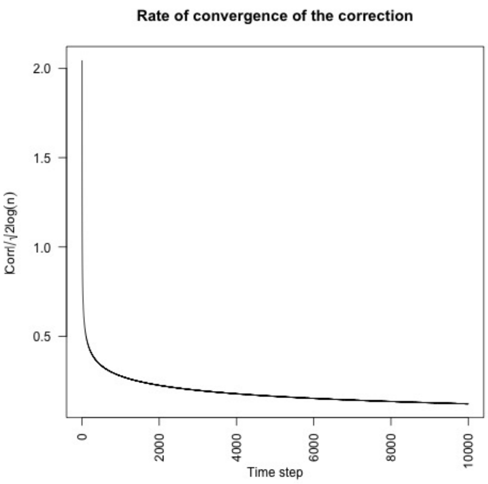


Figure 2.5 – The correction becomes negligible compared to the starting term as n grows large

2.1.2 Sample following a Cauchy distribution

Computing the quantiles Let $X_1, \cdots, X_n$ be i.i.d. RVs ~ *Cauchy(0,1)*. The distribution function is $F_X(t) = \frac{1}{\pi} * \arctan(x) - \frac{1}{2}$. The $\frac{1}{n}$-quantiles satisfy the equation :

$$
\begin{aligned}
& & F_X(q_{\frac{1}{n}}) = 1 - \frac{1}{n} \\
&\iff & \frac{\arctan(q_{\frac{1}{n}})}{\pi} + \frac{1}{2} = 1 - \frac{1}{n} \\
&\iff & \frac{\arctan(q_{\frac{1}{n}})}{\pi} = \frac{2-n}{n} \\
&\iff & q_{\frac{1}{n}} = \tan(\frac{\pi}{2} * \frac{2-n}{n})
\end{aligned}
\tag{2.4}
$$

Figure 2.6 – Below, a realisation of the sequence of maxima for i.i.d. *Cauchy(0,1)* RVs

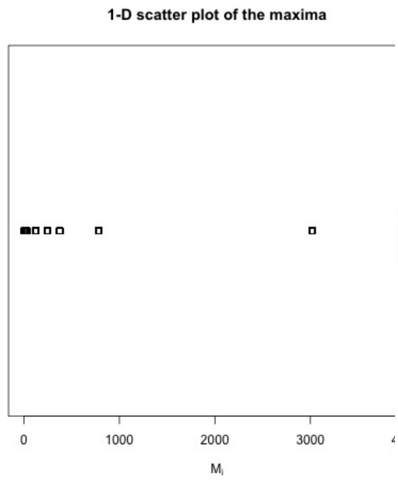


Figure 2.7 – Scatter Plot of the Maxima, n = 10000

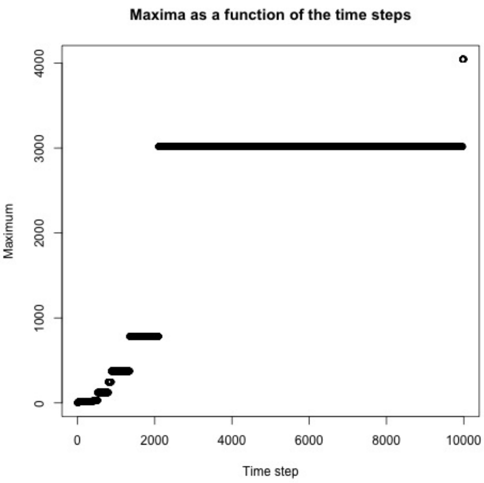


Figure 2.8 – Maxima against the time steps

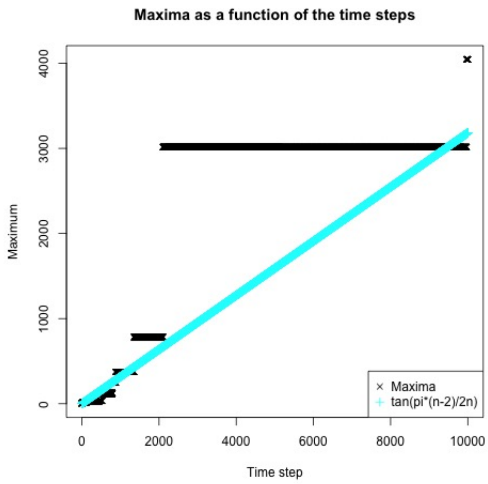


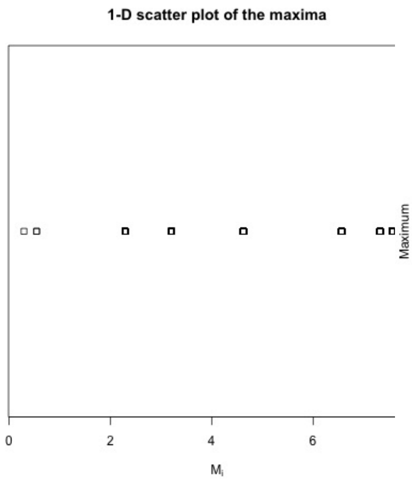
Figure 2.9 – Maxima against the time steps and function n→ $\tan(pi * \frac{2-n}{2*n})$

2.1.3 Sample following an Exponential Distribution

Computing the quantiles Let $X_1, \cdots, X_n$ be i.i.d. RVs ~ *Exponential(λ)*. The distribution function is $F_X(t) = 1 - \exp(-\lambda * t)$. The $\frac{1}{n}$-quantiles satisfy the equation :

$$\begin{aligned} F_X(q_{\frac{1}{n}}) &= 1 - \frac{1}{n} \\ \iff \quad 1 - \exp(-\lambda * q_{\frac{1}{n}}) &= 1 - \frac{1}{n} \\ \iff \quad q_{\frac{1}{n}} &= \frac{1}{\lambda} * \log(n) \end{aligned} \tag{2.5}$$

Figure 2.10 – Below, a realisation of the sequence of maxima for i.i.d. *Exp(1)* RVs



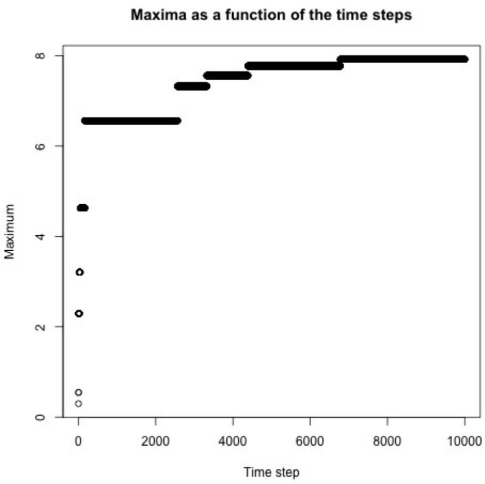


Figure 2.11 – Scatter Plot of the Maxima, n = 10000

Figure 2.12 – Maxima against the time steps

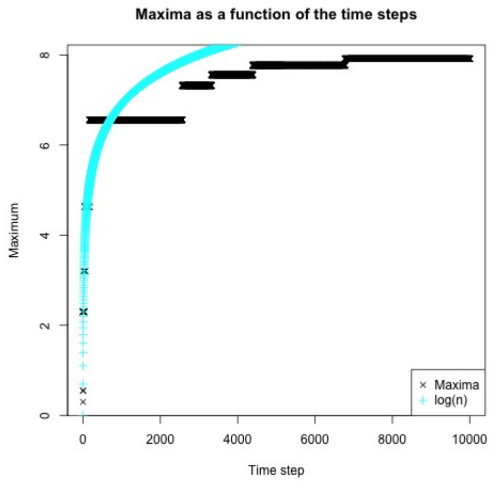


Figure 2.13 – Maxima against the time steps and function n→ $\log(n)$

2.1.4 Why does this work ?

3 Looking into real-world data

A non-numbered chapter...

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

A An appendix

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Bibliography

[Applied Technology Council(1985)] Applied Technology Council. Earthquake damage evaluation data for California. Technical report, Seismic Safety Commission, Applied Technology Council (ATC), California, 1985.

- Modelling Extremal Events, Embrechts, Kluppelberg and Milkosch
- Statistics of Extremes, Theory and Application, Beirlant, Goegebeur, Segers and Teugels
- An Introduction to Statistical Modeling of Extreme Values, Coles
- Introduction to Scientific Programming and Simulation Using R, Jones, Maillardet and Robinson