On the statistical properties and tail risk of violent conflicts

Pasquale Cirillo [a,*], Nassim Nicholas Taleb [b]

[a] *Applied Probability Group, Delft University of Technology, The Netherlands*
[b] *Tandon School of Engineering, New York University, United States*

HIGHLIGHTS

- New methodology to deal with apparently infinite-mean phenomena.
- The new method can be of interest for all researchers dealing with heavy-tailed but bounded random variables.
- New application on war casualties' data, using a novel data set that will be shared upon acceptance.
- Interesting applications in other fields of science.

ARTICLE INFO

ABSTRACT

We examine statistical pictures of violent conflicts over the last 2000 years, providing techniques for dealing with the unreliability of historical data.

We make use of a novel approach to deal with fat-tailed random variables with a remote but nonetheless finite upper bound, by defining a corresponding unbounded *dual distribution* (given that potential war casualties are bounded by the world population). This approach can also be applied to other fields of science where power laws play a role in modeling, like geology, hydrology, statistical physics and finance.

We apply methods from extreme value theory on the dual distribution and derive its tail properties. The dual method allows us to calculate the real tail mean of war casualties, which proves to be considerably larger than the corresponding sample mean for large thresholds, meaning severe underestimation of the tail risks of conflicts from naive observation. We analyze the robustness of our results to errors in historical reports.

We study inter-arrival times between tail events and find that no particular trend can be asserted.

All the statistical pictures obtained are at variance with the prevailing claims about "long peace", namely that violence has been declining over time.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Since the middle of last century, there has been a multidisciplinary interest in wars and armed conflicts, quantified in terms of casualties [1–8]; while other studies have also covered the statistics of terrorism [9–11]. From a statistical point of view, recent contributions have attempted to show that the distribution of war casualties (or terrorist attacks' victims) tends to have heavy tails, characterized by a power law decay [9,12]. Often, the analysis of armed conflicts falls within the broader study of violence [13,4], with the aim to understand whether we as human are more or less violent and aggressive than in

---

the past, and what role institutions played in that respect. Accordingly, the public intellectual arena has witnessed active debates, such as the one between Steven Pinker [4] on one side, and John Gray [14] on the other, concerning the hypothesis that "the long peace" is a statistically established phenomenon, or a mere statistical sampling error, characteristic of heavy-tailed phenomena—the latter case being corroborated by this paper.

Using a new data set containing 565 armed conflicts with more than 3000 casualties over the period 1–2015 AD, we confirm that the distribution of war casualties exhibits a very heavy right-tail. The tail is so heavy that – at first glance – war casualties could represent an infinite-mean phenomenon [15]. But should the distribution of war casualties have an infinite mean, the annihilation of the human species would be just a matter of time, and the sample properties we can compute from data have no meaning at all in terms of statistical inference. In reality, a simple argument allows us to rule out the infiniteness of the mean: no event or series of events can kill more than the whole world population. The support of the distribution of war casualties is thus necessarily bounded, and the true mean cannot be infinite.

Let $[L, H]$ be the support of the distribution of war casualties today. L cannot be smaller than 0, and we can safely fix it at some level $L \gg 0$ to ignore those small events that are not readily definable as armed conflict [7]. As to H, its value cannot be larger than the world population, i.e. 7.2 billion [16] people in 2015 (today's world population can be safely taken as the upper bound, as never before humanity reached similar numbers).

If Y is the random variable representing war casualties, its range of variation is very large and equal to $H - L$. Studying the distribution of Y can be difficult, given its bounded but extremely wide support. Finding the right parametric model among the set of possible ones is hard, while nonparametric approaches are more difficult to interpret from a risk analysis point of view.

Our approach is to transform the data in order to apply the powerful tools of extreme value theory. Since $H < \infty$ we suggest a special log-transformation of the data, different from the others in the literature. This allows us to use tools such as the Generalized Pareto approximation of tails [17,18], simplifying the choice of the model to fit to our data.

Let L and H be respectively the lower and the upper bound of a random variable Y, and define the function

$$\varphi(Y) = L - H \log \left(\frac{H - Y}{H - L} \right). \tag{1}$$

It is easy to verify that

1. $\varphi \in C^{\infty}$,
2. $\varphi^{-1}(\infty) = H$,
3. $\varphi^{-1}(L) = \varphi(L) = L$.

Then $Z = \varphi(Y)$ defines a new random variable with lower bound L and an infinite upper bound. Notice that the transformation induced by $\varphi(\cdot)$ does not depend on any of the parameters of the distribution of Y, and that $\varphi(\cdot)$ is monotone. In what follows, we will call the distributions of Y and Z, respectively the real and the dual distribution.

By studying the tail properties of the dual distribution (the one with an infinite upper bound), using extreme value theory, we will be able to obtain, by reverting to the real distribution, what we call the *shadow* tail mean of war casualties. We will show that this mean can be 3 times larger than the sample mean, but nevertheless finite.

We assume that many observations are missing from the data set (because of under-reported conflicts), and we also take into consideration the fact that war casualties are just non-precise estimates [19], on which historians often have disputes, without anyone's ability to verify the assessments using period sources. For instance, an event that took place in the eighth century, the An Lushan rebellion, is thought to have killed 13 million people, but there is no precise or reliable methodology to allow us to trust that number, which could be conceivably one order of magnitude smaller. For a long time, indeed, an assessment of the drop in population in China was made on the basis of tax census, which might be attributable to a drop in the aftermath of the rebellion in surveyors and tax collectors [20].

Using resampling techniques, we show that our tail estimates are robust to changes in the quality and reliability of data. Our results and conclusions would replicate even if we missed almost a fifth of the data.

Finally, focusing on a more reliable subset covering the last 500 years of data, one cannot observe any specific trend in the number of conflicts, as large wars appear to follow a homogeneous Poisson process. This memorylessness in the data conflicts with the idea that war violence has declined over time, as proposed by Goldstein [2] and Pinker [4].

2. The data

The data set contains 565 events over the period 1–2015 AD; an excerpt is shown in Table 1. Events are generally armed conflicts, such as interstate wars and civil wars, with a few exceptions represented by violence against citizens perpetrated by the bloodiest dictatorships, such as Stalin's and Mao Zedong's regimes. These were included in order to be consistent with previous works about war victims and violence [4,6].

In the words of Wallensteen and Sollenberg [7], "an armed conflict is a contested incompatibility which concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths".

We had to deal with the problem of inconsistency and lack of uniformity in the attribution of casualties by historians. Some events such as the siege of Jerusalem include death from famine, while for other wars only direct military victims