# Bayesian non-parametric inference for multivariate peaks-over-threshold models

Peter Trubey[1*] and Bruno Sansó[1]

[1*]Department of Statistics, University of California Santa Cruz, 1156 High Street, Santa Cruz,95064, CA, USA.

*Corresponding author(s). E-mail(s): ptrubey@ucsc.edu;
Contributing authors: bruno@soe.ucsc.edu;

**Abstract**

We consider a constructive definition of the multivariate Pareto that factorizes the random vector into a radial component and an independent angular component; The former following a univariate Pareto distribution, and the latter defined on the surface of the positive orthant of the unit hypercube. In this paper, we propose a method for inferring the distribution of the angular component. We identify its support as the limit of the positive orthant of the unit $\boldsymbol{p}$–norm spheres, and introduce a projected gamma family of distributions defined as the projection of a vector of independent gamma random variables onto the $\boldsymbol{p}$–norm sphere. This family serves as a building block for a flexible family of distributions obtained as a Dirichlet process mixture of projected gammas. For model assessment, we discuss scoring methods appropriate to distributions on the unit hypercube. In particular, working with the energy score criterion, we develop a kernel metric that produces a proper scoring rule, and present a simulation study to compare different modeling choices using the proposed scoring rules. Finally we apply our approach to describe the dependence structure of the extreme values of the magnitude of the integrated vapor transport (IVT), data describing the flow of atmospheric moisture along the coast of California for the years of 1979 through 2020. We find clear but heterogeneous geographical dependence.

https://www.overleaf.com/project/612e69c665059f5fd7998d90

## 1 Introduction

The statistical analysis of extreme values focuses on inference for rare events that correspond to the tails of probability distributions. As such, it is a key ingredient in the risk assessment of phenomena that can have strong societal impacts like floods, heat waves, high concentration of pollutants, crashes in the financial markets, among others. The fundamental challenge of extreme value theory (EVT) is to use information, collected over limited periods of time, to extrapolate to long time horizons. This sets EVT apart from most of statistical inference, where the focus is on the bulk of the distribution. Extrapolation to the tails of the distributions is possible thanks to theoretical results that give asymptotic descriptions of the probability distributions of extreme values.

Inferential methods for the extreme values of univariate observations are well established and software is widely available (see, for example, **?**). For variables in one dimension the application of EVT methods considers the asymptotic distribution of either the maxima calculated for regular blocks of data,