Stat

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Where Does the Tail Start? Inflection Points and Maximum Curvature as Boundaries

Rafael Cabral[1] | Maria De Iorio[1,2] | Andrea Cremaschi[3]

[1]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore | [2]Institute for Human Development and Potential (IHDP), A*STAR, Singapore | [3]School of Science and Technology, IE University, Madrid, Spain

Correspondence: Andrea Cremaschi (andrea.cremaschi@ie.edu)

Received: 2 February 2025 | **Revised:** 25 April 2025 | **Accepted:** 20 May 2025

Keywords: distribution tails | heavy-tailed distributions | kernel density estimation

ABSTRACT

Understanding the tail behaviour of distributions is crucial in statistical theory. For instance, the tail of a distribution plays a ubiquitous role in extreme value statistics, where it is associated with the likelihood of extreme events. There are several ways to characterize the tail of a distribution based on how the tail function, $\overline{F}(x) = P(X > x)$, behaves when $x \to \infty$. However, for unimodal distributions, where does the core of the distribution end and where does the tail begin? This paper addresses this unresolved question and explores the usage of delimiting points obtained from the derivatives of the density function of continuous random variables, namely, the inflection point and the point of maximum curvature. These points are used to delimit the bulk of the distribution from its tails. We discuss the estimation of these delimiting points and compare them with other measures associated with the tail of a distribution, such as kurtosis and extreme quantiles. We derive the proposed delimiting points for several known distributions and show that it can be a reasonable criterion to define the starting point of the tail of a distribution.

1 | Introduction

In probability theory, the tails of a distribution are commonly studied in the context of the theory of heavy-tailed distributions (Foss et al. 2011; Taleb 2020). Tail behaviour is usually described by the tail function $\overline{F}(x) = P(X > x)$ for $x \to \infty$. For instance, a distribution is heavy-tailed if the tail function is not exponentially bounded, more precisely, $\lim_{x\to\infty} e^{tx}\overline{F}(x) = \infty$, $\forall t > 0$. There are two important subclasses of heavy-tailed distributions: long-tailed (Asmussen 2003), and subexponential distributions (Teugels 1975). For an overview, see Foss et al. (2011) and references therein. These distributions are often used to model the occurrence of extreme events in applied fields such as finance and climate science (Pisarenko and Rodkin 2010; Longin 2016; Cabral et al. 2023a). Also, in robust statistics, heavy-tailed distributions can provide more robust inferences since they can reduce the sensitivity of the linear regression estimates with regards to outliers present in the data (Huber 2004; Cabral et al. 2023b). The tails of the distribution are also important to describe the behaviour of shrinkage priors in sparse regression (Carvalho et al. 2010).

Existing analyses of distribution tails primarily focus on their asymptotic behaviour as x approaches infinity, often overlooking what happens before infinity. Also, the tail is not a precisely defined interval of the probability density function (pdf), in the sense that there is no specific point where the bulk of a distribution ends and the tail begins. To our knowledge, there has yet to be a concrete discussion in the literature regarding the definition of a delimiting point for the left tail, t_l, or the right tail, t_r. This paper presents automatic approaches for selecting these delimiting points. We will concentrate our attention on continuous

random variables with support on $\mathbb{R}$ or $[C, +\infty]$ with $C \in \mathbb{R}$, thus also including $\mathbb{R}^+$ when $C = 0$, and characterized by a unimodal pdf, where the delimiting points will be used to define the bulk or modal region, $[t_l, t_r]$, the proper tail interval $[t_r, \infty]$, and finally the left tail interval $[C, t_l[$ of the pdf. These intervals can also be applied to other functions that describe the distribution, such as the tail function $\overline{F}(x)$ or the hazard function $h(x) = f(x)/\overline{F}(x)$.

While the terms 'bulk' and 'tail' have not been precisely defined, they are often used informally to describe different distribution regions and convey an intuitive understanding of where the probabilities are concentrated. For unimodal distributions with support on the real line, the bulk or modal region of the pdf colloquially refers to a high-density interval centered around the mode. The bulk contains the values that are more likely to be observed. As we move towards the tails of the distribution, the probability density decreases, indicating that values in the tails are less likely to occur (see Figure 1).

The 'bulk' and 'tail' regions often correspond to different behaviours of the second derivative of the density function. For example, in the case of the Cauchy distribution illustrated in Figure 1, the density function is concave for $|x| < 1/\sqrt{3}$ and behaves like $(1-x^2)/\pi + \mathcal{O}(x^3)$ near the origin. On the other hand, for $|x| > 1/\sqrt{3}$, it becomes convex and behaves like $1/(\pi x^2) + \mathcal{O}(x^{-3})$ as x approaches infinity. The inflection points $x \pm 1/\sqrt{3}$ mark the change from concavity to convexity.

In this paper, we explore different methods that can be used to define modal regions and the tail's starting point based on the pdf's second derivative. The inflection point of the pdf is a natural candidate for many unimodal distributions since it marks the pdf's change from concavity to convexity. This will yield similar modal regions as in Duong et al. (2008), which focuses on regions where the pdf is significantly concave. Here, we propose using the point of the maximum curvature and contrast it with other measures related to heavy-tailedness, such as the kurtosis and extreme quantiles for commonly used distributions. A better understanding of the derivatives of density functions not only advances statistical theory but also opens doors for new methodologies for data analysis. For instance, an adequate identification of modal regions is useful in applications such as nonparametric clustering and bump hunting, as discussed in Siloko et al. (2019) and Chacón and Duong (2013).

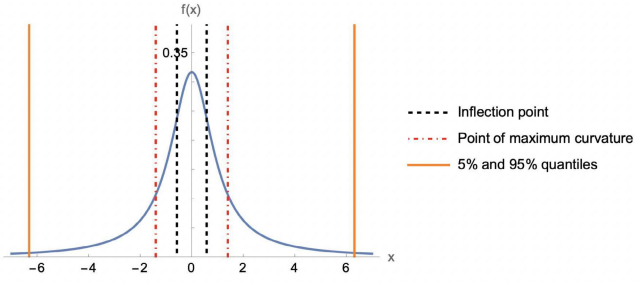


FIGURE 1 | Density function of Cauchy distribution (blue line). The inflection point, point of maximum curvature, and (5% and 95%) quantiles are marked as vertical dashed, dotted, and continuous lines, respectively.

In statistical data analysis, Duong et al. (2008) develop a test for the null hypothesis that the hessian of the pdf is positive definite to identify significant concave regions. The method aims to identify multimodality and clusters in the data and is used to identify spatial hotspots in earthquake occurrences and high-density spots for large biological cells in flow cytometry studies. Another area where finding a transition point between the bulk and tail of the distribution is useful is in extreme value theory. It is common to exclude all but the most extreme observations when interest focuses on estimating only tail features, for example, estimating the tail index α[1] or extrapolating to high quantiles from limited observed data. The Pickands-Balkema-de Haan Theorem (Balkema and De Haan 1974; Pickands 1975) justifies the peaks-over-threshold estimation method, which involves fitting a generalized Pareto distribution (GPD) to observations surpassing a specified threshold. Data-driven threshold selection methods are thus crucial and commonly rely on graphical diagnostic tools or automatic techniques (Scarrott and MacDonald 2012). In general, single-threshold selection involves a bias-variance trade-off, where lower thresholds increase model misspecification bias and higher thresholds increase estimation uncertainty, and while many approaches focus on sensitivity analysis or selecting a suitable threshold, Northrop et al. (2017) address such trade-off using cross-validation to optimize threshold selection for improved prediction of future extreme events. Murphy et al. (2024) propose the expected quantile discrepancy (EQD) method, an adaptation of the approach by Varty et al. (2021), to estimate a threshold for which the sample excesses, arising from iid and nonmissing realizations of a continuous random variable, are most consistent with a GPD model. The EQD uses a QQ-plot-based metric to approximate the integrated absolute error between the quantiles of the model and the data-generating process. However, determining the optimal threshold presents challenges, for instance, due to potentially multiple transition regimes to the tail. Various approaches have emerged to estimate the entire density function by employing a mixture model, which considers a separate model for the bulk and a GPD model for the tail, with examples including those proposed by Tancredi et al. (2006), MacDonald et al. (2011), and do Nascimento et al. (2012). However, these methods often rely on heuristic techniques to find the threshold with limited mathematical analysis support.

The paper is organized as follows. In Section 2, we define the curvature of a pdf, the delimiting points between the bulk and tail of the distribution that can be constructed from it, and examine their properties, such as their existence and how they are related. Section 4 addresses the estimation of these delimiting points from the data, where we provide sample versions of these points and prove their uniform consistency. In Section 3, we derive the delimiting points for several known distributions. Lastly, in Section 7, we discuss the main results.

2 | Curvature

The concept of curvature is fundamental in mathematics and is widely explored in various fields. In this section, we define the curvature and, based on it, construct useful delimiting points between the bulk and tail of the distribution, namely the point of inflection and maximum curvature. We also present some

properties of these delimiting points and examine their existence and how they are related to one another.

Consider the parametric representation $\gamma(x) = (z(x), y(x))$ of a plane curve, assumed to be twice differentiable. We assume the derivative $d\gamma/dx$ is well-defined, differentiable, and not identical to the zero vector across the parametrization domain. Utilizing this parametrization, the signed curvature can be expressed as

$$\kappa(x) = \frac{z'(x)y''(x) - y'(x)z''(x)}{\left((z'(x))^2 + (y'(x))^2\right)^{3/2}}$$

with primes denoting derivatives with respect to x. Intuitively, the curvature is the amount by which a curve deviates from being a straight line. Our interest, pertaining to the establishment of a probability density function tail starting point, lies in the graph of a function $y = f(x)$, where $f(x)$ is the density function of a given random variable X. This is a specific instance of a parametrized curve defined as

$$z(x) = x$$
$$y(x) = f(x)$$

Given that the first and second derivatives of z with respect to x are 1 and 0, respectively, the above formula can be simplified to

$$\kappa(x) = \frac{f''(x)}{\left(1 + f'(x)^2\right)^{3/2}} \tag{1}$$

2.1 | Point of Inflection and Maximum Curvature

The graph of the differentiable function has an inflection point at $(x, f(x))$ if and only if its first derivative f' has an isolated extremum at x. In our context, the inflection point is given by $\text{PInf} = \arg\max_x |f'(x)|$ and can also be found by computing the roots of the second derivative of the pdf. On the other hand, the point of maximum curvature is the point where the curvature $\kappa(x)$ in (1) achieves its highest value:

$$\text{PMCurv} = \arg\max_x \kappa(x)$$

In general, there is no closed-form expression for PInf and PMCurv, and they need to be computed using numerical optimization algorithms.

To simplify calculations, the curvature in (1) can be approximated by the second derivative of the pdf when the derivative in the denominator is much smaller than 1 in absolute value. Indeed, we have

$$\kappa(x) = f''(x)\left(1 + O\left(f'(x)^2\right)\right)$$

In the context of probability distributions, this can happen when we apply a scale transformation $X \to \sigma X$ and σ is sufficiently large. If $f(x)$ denotes the probability density function of X, then the distribution of σX is $f_\sigma(x) = f(x/\sigma)/\sigma$, and $f'_\sigma(x) = {}' (x/\sigma)/\sigma^2$ which goes to 0 when $\sigma \to \infty$. Intuitively, the densities become flatter when σ increases, with the first derivative going to zero. In this case, the curvature of $f(x)$ is $\kappa(x) \approx f''(x)$ and the point of maximum curvature can be approximated by the point of maximum convexity (PMConv), which is defined as

$$\text{PMCurv} \approx \text{PMConv} = \arg\max_x f''(x) \tag{2}$$

We now study the behaviour of PInf_X, PMConv_X and PMCurv_X of a random variable X when applying a location-scale transformation, that is, $X \to \sigma X + \mu$. We have that $\text{PInf}_{\mu+\sigma X} = \mu + \sigma\text{PInf}$, $\text{PConv}_{\mu+\sigma X} = \mu + \sigma\text{PMConv}_X$ and $\text{PMCurv}_{X+\mu} = \text{PMCurv}_X + \mu$. On the other hand, in general, $\text{PMCurv}_{\sigma X} \neq \sigma\text{PMCurv}_X$. However, under the approximation in Equation (2), which holds for large σ, we have

$$\text{PMCurv}_{\sigma X} \approx \text{PMConv}_{\sigma X} = \sigma\text{PMConv}_X \tag{3}$$

In Figure 2, we show the PMCurv and PMConv for several continuous random variables. In Sections 4 and 3, we consider the approximation PMCurv ≈ PMConv.

Note that for unimodal distributions with mode at θ, $f''(x)$ typically possess two local maximums, one at $x < \theta$ relating to the left tail, PMConv_l, and another at $x > \theta$, relating to the right tail, PMConv_r (see Figure 1). For symmetric unimodal distributions with zero mean $\text{PMConv}_l = -\text{PMConv}_r$, but generally this is not the case (for example, see the log-normal distribution in Section 3.1). Thus, it is useful to define

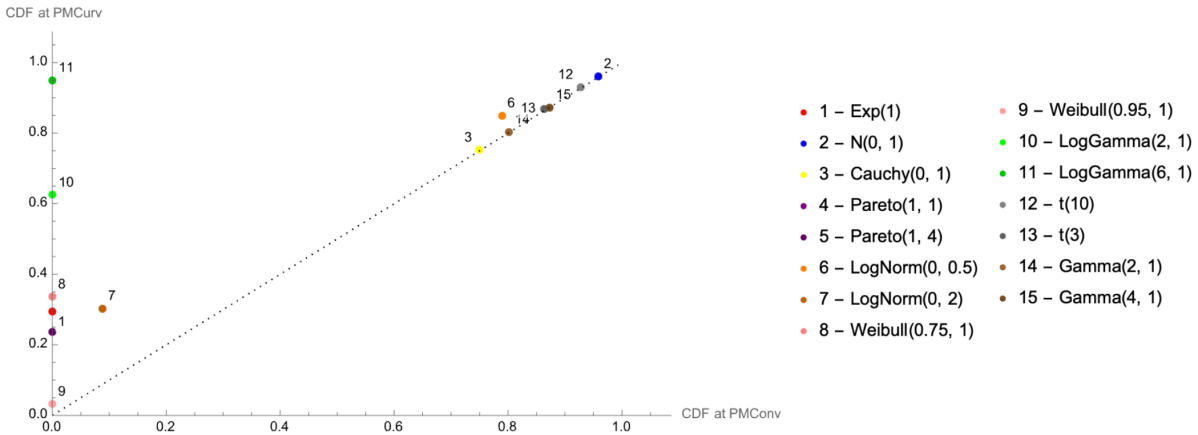


FIGURE 2 | Cdf of PMCurv_r against cdf of PConv_r for several distributions.

$$\text{PMConv}_l = \arg\max_{x<\theta} f''(x) \quad \text{and} \quad \text{PMConv}_r = \arg\max_{x>\theta} f''(x) \tag{4}$$

where we assume f'' has an isolated maximum and root on each side. Similarly, we define the inflection point associated with the left and right tail:

$$\text{PInf}_l = \arg\max_{x<\theta} |f'(x)| \quad \text{and} \quad \text{PInf}_r = \arg\max_{x>\theta} |f'(x)| \tag{5}$$

Moreover, if $\text{PMConv}_r = 0$ for a random variable X with support on $\mathbb{R}^+$, then for large enough σ, $\text{PMCurv}_{\sigma X}$ is also equal to 0 (see equation 3). This happens, for instance, for the exponential, Pareto, Weibull (for rate parameter $\beta = 1$), and log-Gamma distributions (see Figure 2). The density functions of these distributions have a singularity at 0, as well as their second derivative (thus $\text{PMConv}_r = 0$). Furthermore, the second derivative is always positive, and thus there are no inflection points either. For such distributions, which have a mode at 0, the delimiting points we are considering are not useful in the sense of not providing a practical delimiting point between the modal region and the tail of the distribution.

Finally, it is also useful to work with $F(\text{PMConv}_r)$ and $F(\text{PInf}_r)$, the cdf of the distribution evaluated at PMConv_r and PInf_r (and likewise for PMConv_l and PInf_l). Unlike PMConv_r, which changes under a scale and location transformation, we can use $F(\text{PMConv}_r)$ as a measure that is invariant under scale and location transformations, which facilitates interpretation when comparing different distributions (see Figure 2 and 3). For instance, this measure is about 0.9584 for the standard normal distribution, while for the Cauchy distribution, it is equal to 0.75. As we will see in Section 3, $F(\text{PInf}_r)$ and $F(\text{PConv}_r)$ tend to decrease as the distribution becomes more heavy-tailed for several families, such as the Student's t-distribution, which encompasses the Gaussian and Cauchy distributions. We also observe that for the Gaussian distribution, PConv_l and PConv_r are very close to the 5% and 95% quantiles, which are commonly used to define extreme values or outliers.

We now discuss the existence of the delimiting points introduced so far and prove their existence in the case of unimodal distributions.

Theorem 1. *Let $f(x)$ be the unimodal density function of a continuous random variable with support on the real line, with mode θ, and let $f^{(d)}(x)$ be uniformly continuous for $d = 0, 1, 2$. Then, $f(x)$ has at least one inflection point and one point of maximum convexity for $x > \theta$*, and similarly for $x < \theta$.

Proof. The proof is given in Appendix S1. □

For unimodal distributions with support on $[C, \infty]$, $C \in \mathbb{R}$, and mode in $\theta > C$, the same argument can be used to show that an inflection point and PMConv exist for $x > \theta$. When the unique mode is at C, the boundary of the support, we further require that $f''(C) < 0$; otherwise, these delimiting points may not exist for $x > C$. An example is the exponential distribution, which has a positive second derivative for $x \geq 0$, with a maximum at $x = 0$, and therefore, no inflection point, and the point of maximum convexity is at 0. The following proposition indicates that the modal region $]\text{PInf}_l, \text{PInf}_r[$ is smaller than $]\text{PMConv}_l, \text{PMConv}_r[$.

Proposition 1. *Under the conditions of Theorem 1, if, to the right of the mode, there is only one inflection point,* PInf_r*, and one point of maximum convexity* PMConv_r*, then* $\text{PMConv}_r > \text{PInf}_r$*. Similarly, at the left of the mode,* $\text{PMConv}_l < \text{PInf}_l$.

Proof. The proof is given in Appendix S1. □

3 | Examples

In this section, we investigate the delimiting points of several known distributions based on the methods described in Section 2. Following the discussion in Section 2, we consider distributions with support on $\mathbb{R}^+$ with mode larger than 0 (log-normal distribution), with mode at 0 (exponential distribution) and several unimodal distributions with support on

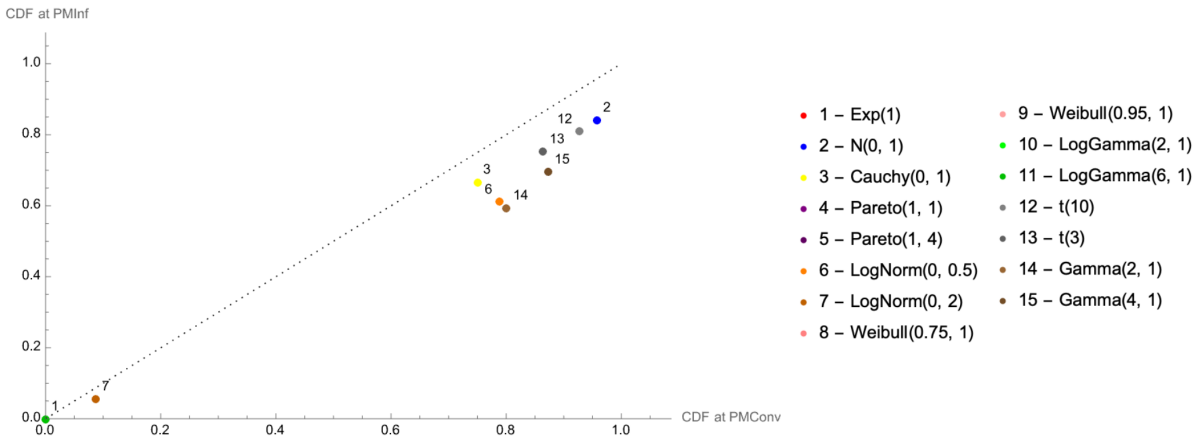


FIGURE 3 | Cdf at PInf_r against cdf at PConv_r for several distributions. Many of the distributions do not have PInf_r, and for those, we omit the results.

$\mathbb{R}$ (Gaussian, Student's t-distribution and Skew-t distribution). We compare these delimiting points with measures of heavy-tailedness and asymmetry, namely the kurtosis and extreme quantiles.

3.1 | Log-Normal Distribution

Figure 4 shows the density function of the log-normal distribution, its second derivative, and the delimiting points. The point of maximum convexity for the left tail is $\text{PMConv}_l = e^{\mu - 2\sigma^2 - \sigma\sqrt{3+\sigma^2}}$ and for the right tail is $\text{PMConv}_r = e^{\mu - 2\sigma^2 + \sigma\sqrt{3+\sigma^2}}$. These delimiting points are used to define the modal region $[\text{PMConv}_l, \text{PMConv}_r]$, the left tail interval $[0, \text{PMConv}_l[$ and the right tail interval $[\text{PMConv}_r, \infty]$. Likewise, the modal region based on the inflection points is given by

$$\left[e^{\frac{1}{2}\left(2\mu - 3\sigma^2 - \sigma\sqrt{\sigma^2+4}\right)}, e^{\frac{1}{2}\left(2\mu - 3\sigma^2 + \sigma\sqrt{\sigma^2+4}\right)}\right]$$

The larger σ, the slower the tails of the density function decay, and the higher the kurtosis is ($\text{Kurt} = 3e^{2\sigma^2} + 2e^{3\sigma^2} + e^{4\sigma^2} - 3$). On the other hand, the modal regions defined by the inflection point and point of maximum convexity shrink, and the delimiting points for the tails get closer to the mode. Moreover, the 95% quantile is given by $e^{\mu + 1.64485\sigma}$, and thus, if one uses the quantiles, the modal region widens as we increase σ, having the opposite behaviour as the modal region defined by the previous delimiting points (see Figure 5).

3.2 | Exponential Distribution

As shown in Figure 6, the density of the exponential distribution has a singularity at 0 and no inflection points. For this distribution, $\text{PMConv} = 0$ since the second derivative of the pdf has a maximum at 0, but $\text{PMCurv} = \log(2\lambda^6)/(2\lambda)$ for $\lambda > 2^{-1/6}$, and is 0 otherwise, where λ is the rate parameter. Thus, it would be more appropriate to define the modal region $[0, t_r]$ and right tail $[t_r, \infty]$ in terms of the quantiles (see Figure 7).

3.3 | Gaussian Distribution

Figure 8 shows the density of the Gaussian distribution, where the inflection points are exactly one standard deviation from the mean: $\mu \pm \sigma$ (PMInf_r corresponds to the quantile 0.841). Now, the points of maximum convexity are located at $\sqrt{3}\sigma$ standard deviations from the mean: $\mu \pm \sqrt{3}\sigma$ (PMConv_r corresponds to quantile 0.958). There is no closed-form expression for the PMCurv, although, for large σ, it tends to $\text{PMConv} = \mu \pm \sqrt{3}\sigma$, as can be seen in Figure 9.

3.4 | Student's t distribution

Figure 10 shows the density function and its second derivative of a Student's t-distribution with 3 degrees of freedom. The points of inflection and maximum convexity are at $\pm\sqrt{\nu}/\sqrt{\nu+2}$ and $\pm\sqrt{3\nu}/\sqrt{\nu+2}$, respectively. The tails decay according to $x^{-\nu-1}$ and the kurtosis is $3 + 6/(\nu - 4)$, when $\nu > 4$. As ν decreases, the tail decays more slowly, and the kurtosis increases. Also, as ν decreases, the modal regions defined by $[\text{PInf}_l, \text{PInf}_r]$ and $[\text{PMConv}_l, \text{PMConv}_r]$ shrink, and the delimiting points for the tails get closer to the mode. On the other hand, the modal region defined by the 5% and 95% quantiles widens as the distributions become more heavy-tailed (see Figure 11). The previous inverse relationship between kurtosis and PInf_r and PMConv_r is present for many unimodal distributions, although an exact relationship is not straightforward to derive. For the Cauchy distribution, PMInf_r and PMConv_r correspond to the quantiles 2/3 and 3/4, respectively.

3.5 | Skew-t distribution

Figure 12 shows the density function and its second derivative of a skew-t distribution with parameters $\nu = 3$ and $s = 10$. Branco and Dey (2001) and Azzalini and Capitanio (2003) define the density function of a skew-t distribution as:

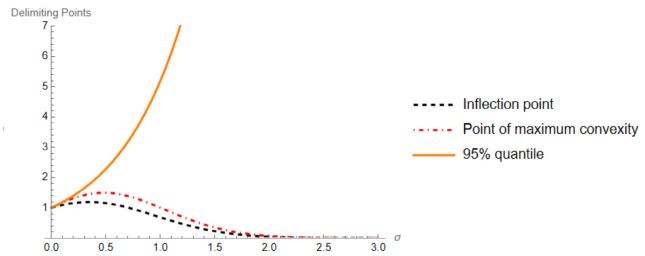


FIGURE 5 | Delimiting points for the right tail as a function of σ for a log-normal distribution with parameters $\mu = 0$ and σ.

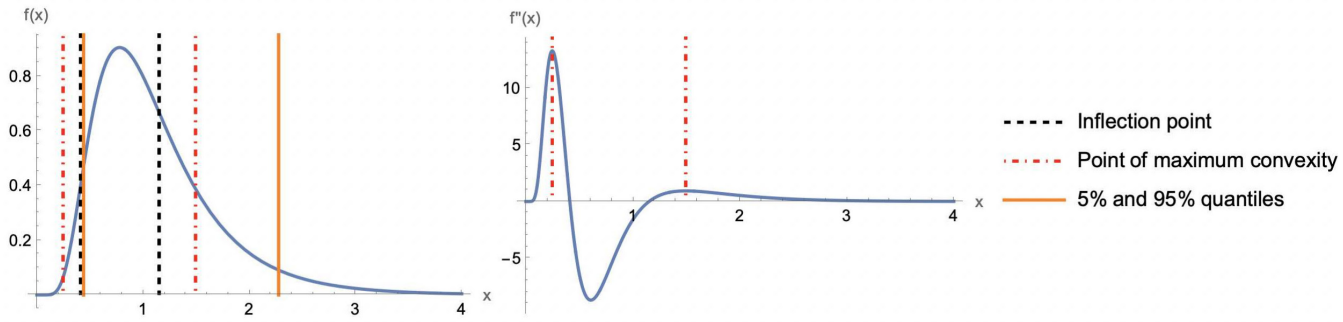


FIGURE 4 | Density function and its second derivative of a log-normal distribution with parameters $\mu = 0$ and $\sigma = 0.5$. The dashed, dotted and continuous vertical lines are drawn in correspondence to the PInf, PMConv, and the 5% and 95% quantiles, respectively.

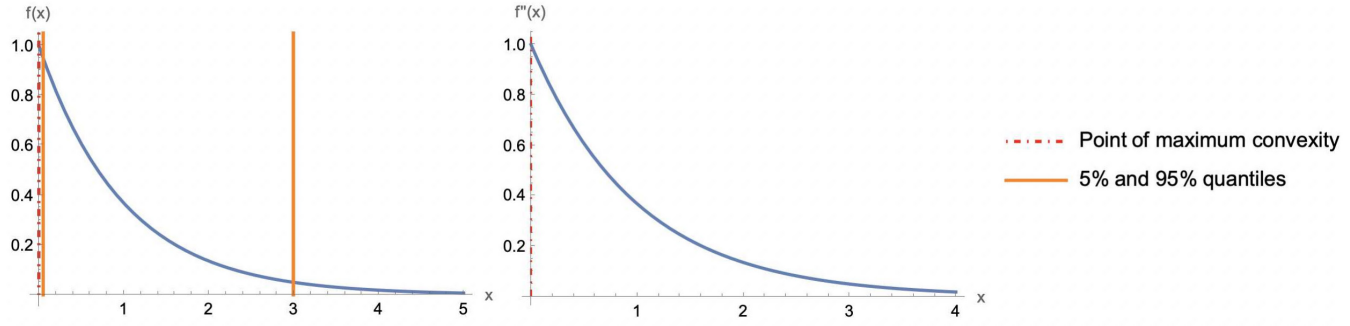


FIGURE 6 | Density function and its second derivative of an exponential distribution with rate parameter 1. The dotted and continuous vertical lines are drawn in correspondence with the PMConv and the 5% and 95% quantiles, respectively.

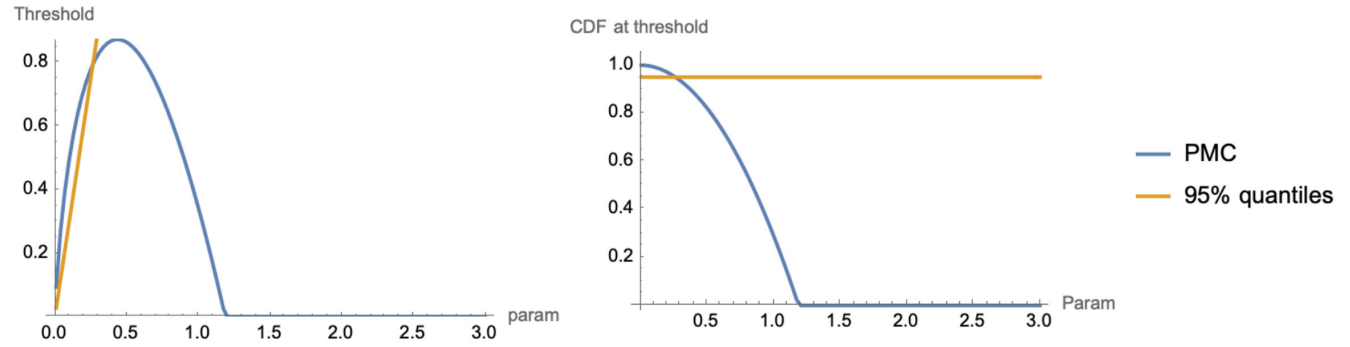


FIGURE 7 | PMCurv and 95% quantile of an exponential distribution with varying scale parameter.

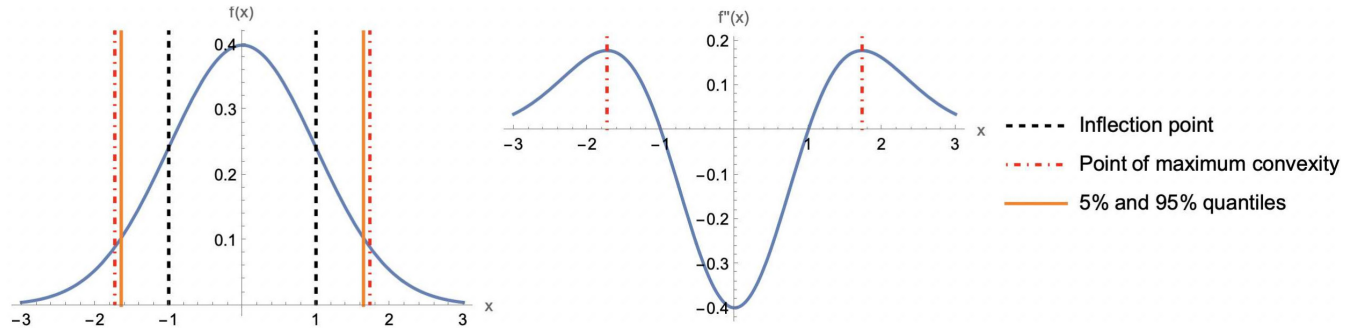


FIGURE 8 | Density function and its second derivative of a standard normal distribution. The dashed, dotted and continuous vertical lines are drawn in correspondence with the PInf, PMConv, and the 5% and 95% quantiles, respectively.

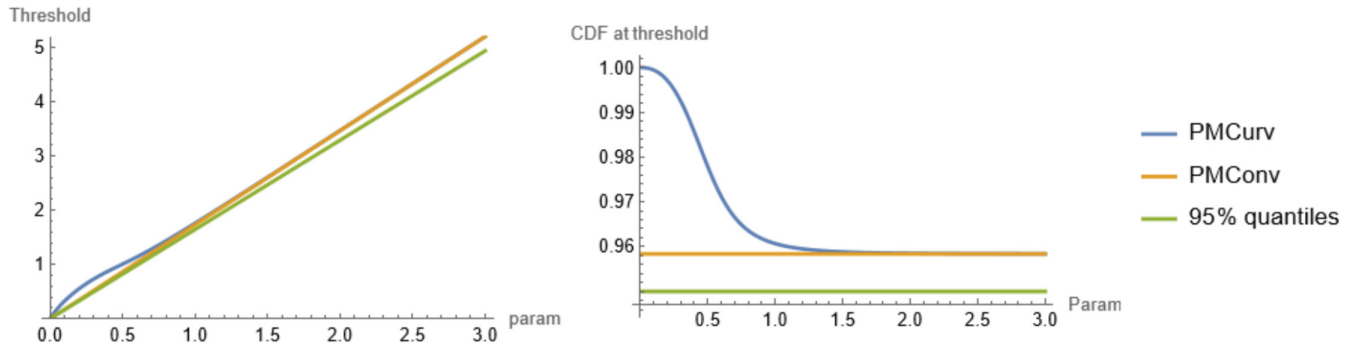


FIGURE 9 | PMCurv, PMConv and 95% quantile of a Gaussian distribution with varying scale parameter.

$$\frac{2}{\sigma}t(z;\nu)T\left(sz\sqrt{\frac{\nu+1}{\nu+z^2}};\nu+1\right),\quad z=\frac{x-\mu}{\sigma}$$

where $t(x;\nu)$ and $T(x;\nu)$ are the pdf and cdf of a symmetric Student's t-distribution with ν degrees of freedom and scale parameter 1, where the parameter s regulates the skewness. We compute the inflection points and point of maximum convexity numerically for $\mu=0$, $\sigma=1$, $s\in[-10,10]$, and $\nu\in[1,20]$. Figures 13 and 14illustrate that as ν and s increase, PInf_r and PMConv_r move towards the center of the distribution. Interestingly, heightened levels of skewness bring the previous points closer to the mode. This behaviour is also found for the normal inverse Gaussian distribution, which is a skewed and leptokurtic distribution (Barndorff-Nielsen 1997).

4 | Estimation

The inflection points and points of maximum convexity are related to the extremums of the first and second derivatives of the pdf, respectively. Therefore, we need to study:

$$\theta^d=\arg\max_{-\infty<x<\infty}f^{(d)}(x) \tag{6}$$

We present an estimator for these maximizers and discuss their uniform consistency. Finally, we present a simulation study and highlight some difficulties in estimating PMCurv and PInf.

For the inflection point, we are also interested in computing $\arg\min_x f'(x)$, and it is straightforward to extend the arguments below to this case. Also, for the unimodal distributions of Section 3, there are two local maximums for the first and second derivatives that we are interested in, one pertaining to the left tail and the other to the right tail. In this case, we can find unique maximizers for each tail by constraining the optimization region of (6) to be at the right or left of the mode.

The estimation of θ^d is ultimately related to the estimation of the probability density function $f(x)$ and its derivatives. Let $X_1,\ \ldots,X_n$ be independent and identically distributed with distribution function $F(x)=\int_{-\infty}^{x}f(u)du$. Davis et al. (2011) considered nonparametric estimators of the form

$$f_n(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}K\left(\frac{x-X_i}{h_n}\right)$$

where the kernel function $K(x)$ is a probability density function and h_n is the bandwidth which goes to 0 as n increases. Usually, $K(x)$ satisfies the following conditions:

$$\int K(x)dx=1,\ \int xK(x)dx=0\ \text{and}\ \int x^2K(x)dx\neq 0$$

Consider now the derivatives of $f_n(x)$:

$$f_n^{(d)}(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n^{d+1}}K^{(d)}\left(\frac{x-X_i}{h_n}\right) \tag{7}$$

From Equation (7), it is clear that when using a kernel estimator for the dth derivative of a function, two key choices must be made: the kernel function K and the smoothing parameter or bandwidth h_n. As noted in Guidoum (2015), the choice of the kernel K is generally less critical. In contrast, determining an appropriate bandwidth h_n is a more challenging task, as it strongly impacts the shape of the resulting estimator (Jones et al. 1996). A small bandwidth tends to produce an undersmoothed estimator with high variance, while a large bandwidth results in an overly smooth estimator that may deviate considerably from the true underlying function. Guidelines on the choice of h_n when estimating derivatives via kernel density estimation are provided in Härdle et al. (1990) and Politis et al. (2015). Asymptotic properties of this estimator, including uniform consistency and asymptotic normality, are derived in Parzen (1962), Schuster (1969) and Bhattacharya (1967).

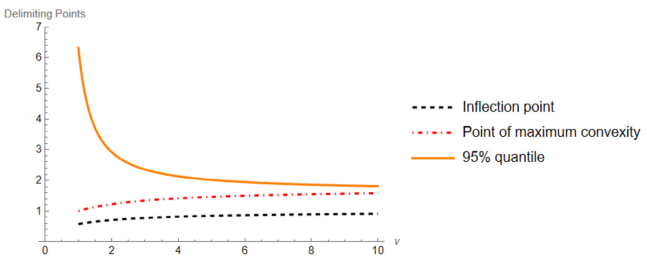


FIGURE 11 | Delimiting points for the right tail as a function of the degrees of freedom ν of a Student's t-distribution.

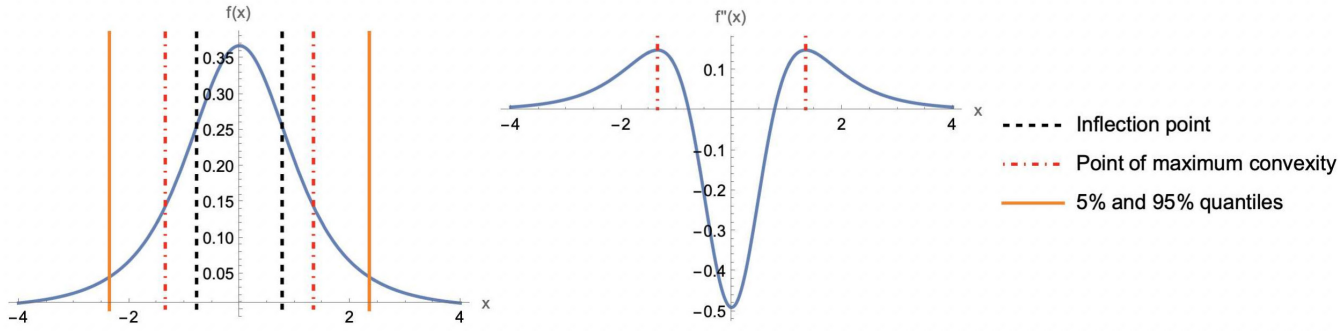


FIGURE 10 | Density function and its second derivative of a Student's t-distribution with 3 degrees of freedom. The dashed, dotted and continuous vertical lines are the PInf, PMConv, and the 5% and 95% quantiles, respectively.

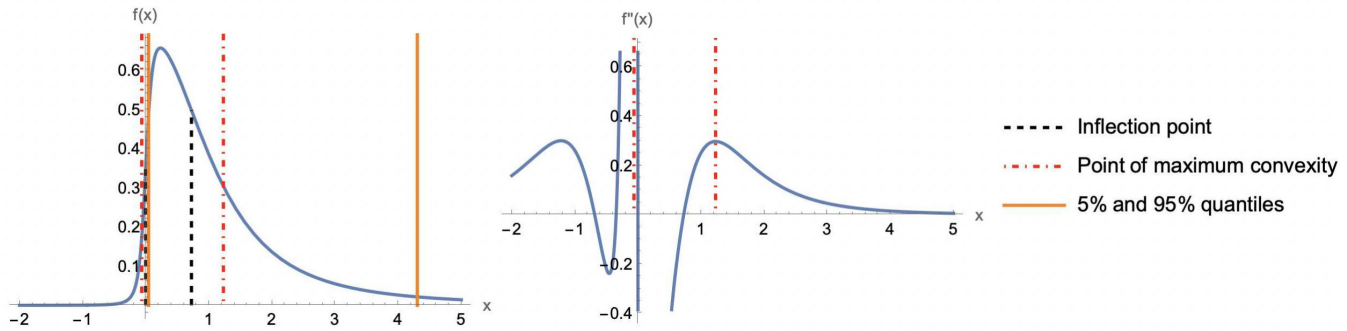


FIGURE 12 | Density function and its second derivative of a Skew-t distribution with parameters $\nu = 3$ and $s = 10$. The dashed, dotted and continuous vertical lines are drawn in correspondence with the PInf, PMConv, and the 5% and 95% quantiles, respectively.

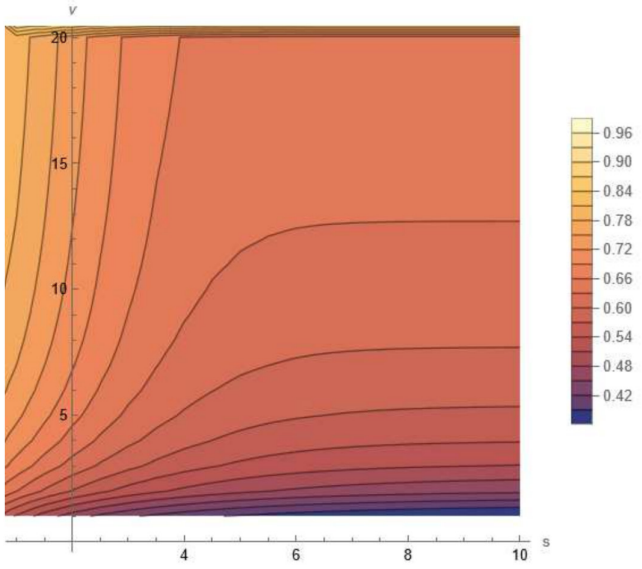


FIGURE 13 | Quantile corresponding to PInf_r of a skew-t with parameters ν and s.

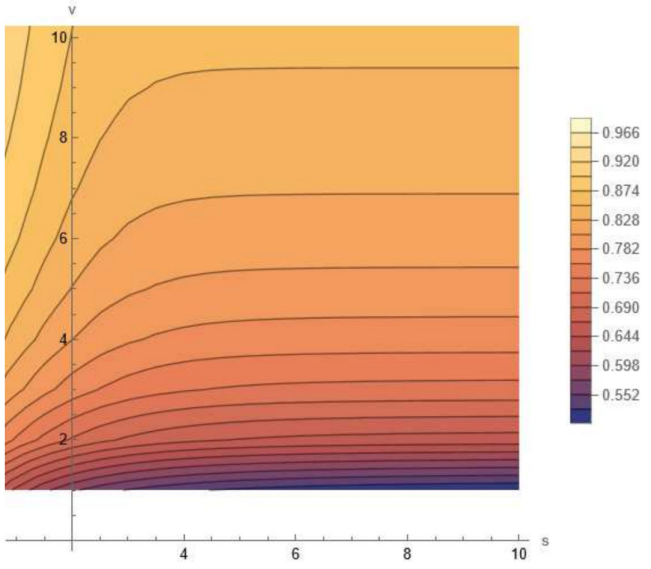


FIGURE 14 | Quantile corresponding to PConv_r of a skew-t with parameters ν and s.

We assume $\int |u|K(u)du$ is finite and $K^{(d)}(x)$ is a continuous function of bounded variation. These conditions are satisfied, for instance, when $K(x)$ is the density function of a standard normal distribution. We further assume that $f^{(d)}(x)$ is uniformly continuous and possesses a maximizer θ^d defined by

$$f(\theta^d) = \max_{-\infty < x < \infty} f^{(d)}(x)$$

and that this maximizer is unique. An estimator for θ^d is given by

$$\theta_n^d = \arg\max_{-\infty < x < \infty} f_n^{(d)}(x)$$

and in the following theorem, we show the consistency of θ_n^d as an estimator of θ^d.

Theorem 2. Consistency of θ_n^d *as an estimator of* θ^d. *If* h_n *is a function of* n *satisfying:*

$$n \to \infty \lim nh_n^{2d+2} = \infty$$

then for every $\epsilon > 0$

$$P(|\theta_n^d - \theta^d| > \epsilon) \to 0 \quad \text{as } n \to \infty$$

Proof. The proof is given in Appendix S1. □

5 | Simulation Study

We utilize the Gaussian kernel because it has derivatives of all orders and simplifies the required mathematical computations. The derivatives are expressed as $K^{(d)}(x) = (-1)^d H_d(x)K(x)$, where $H_d(x)$ represents the dth Hermite polynomial. The initial five Hermite polynomials are denoted as $H_0(x) = 1, H_1(x) = x, H_2(x) = x^2 - 1, H_3(x) = x^3 - 3x$, and $H_4(x) = x^4 - 6x^2 + 3$. Therefore, an estimator for the dth derivative of the density function is given by

$$f_n^{(d)}(x) = \frac{(-1)^d}{\sqrt{2\pi}nh^{d+1}} \sum_{i=1}^{n} H_d\left(\frac{x - X_i}{h}\right) \exp\left(-\frac{1}{2}\left(\frac{x - X_i}{h}\right)^2\right)$$

TABLE 1 | MSE and MAE for the estimated inflection points considering n simulated observations from a Student's t-distribution with ν degrees of freedom.

	MSE			MAE		
	$\nu = 1$	$\nu = 5$	$\nu = 100$	$\nu = 1$	$\nu = 5$	$\nu = 100$
$n = 100$	0.429	0.204	0.106	0.609	0.397	0.273
$n = 500$	0.207	0.090	0.045	0.439	0.270	0.180
$n = 2000$	0.116	0.044	0.023	0.332	0.188	0.128

TABLE 2 | MSE and MAE for the estimated points of maximum convexity considering n simulated observations from a Student's t-distribution with ν degrees of freedom.

	MSE			MAE		
	$\nu = 1$	$\nu = 5$	$\nu = 100$	$\nu = 1$	$\nu = 5$	$\nu = 100$
$n = 100$	1.139	0.599	0.312	0.920	0.679	0.488
$n = 500$	0.554	0.321	0.160	0.665	0.501	0.357
$n = 2000$	0.332	0.175	0.087	0.520	0.376	0.266

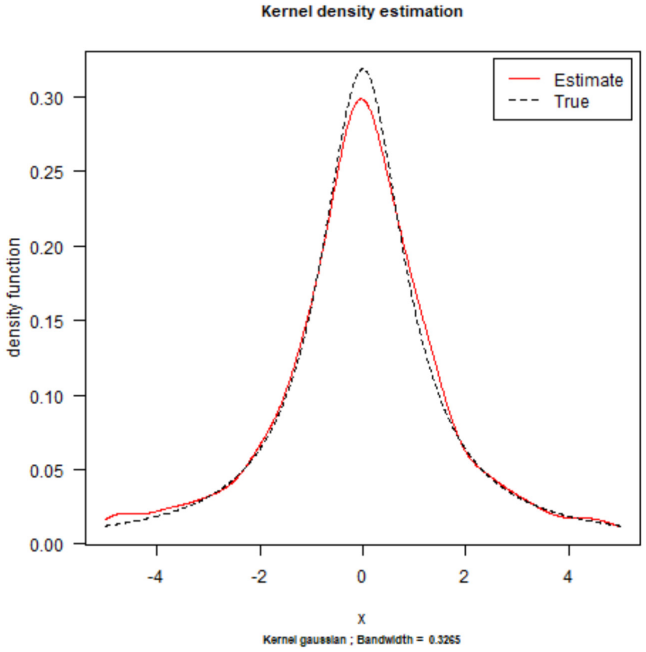


FIGURE 15 | Density function of a Cauchy distribution estimated from 2000 simulated observations.

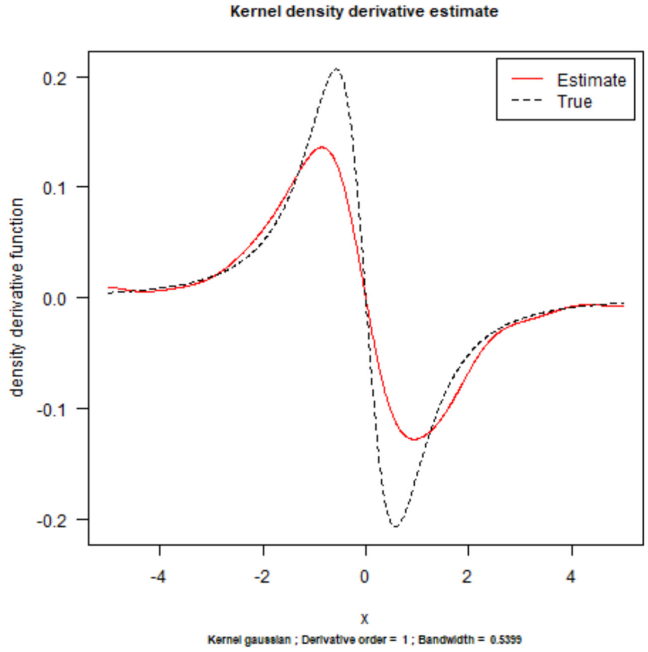


FIGURE 16 | First derivative of the density function of a Cauchy distribution estimated from 2000 simulated observations.

We chose h_n to minimize the asymptotic mean integrated squared error (AMISE), as detailed in Guidoum (2020) and Siloko et al. (2019). Namely, we have

$$h_n^d = \left[\frac{(2d+1)R\left(K^{(d)}\right)}{\mu_2(K)^2 R\left(f^{(d+2)}\right)}\right]^{\frac{1}{2d+5}} \times n^{-\frac{1}{2d+5}} \tag{8}$$

where $R(g) = \int g(x)^2 dx$. There are several estimators available for $R\left(f^{(d+2)}\right)$, such as those discussed in Guidoum (2020). However, in this simulation study, we utilize the true value $R\left(f^{(d+2)}\right)$ because the density function f of the simulated data is known.

In the following subsections, we present examples of the proposed approach to estimate the points of inflection and of maximum curvature, for all the distributions introduced in Section 3, with exception of the skew-t distribution for which an closed-form expression of these points is not available.

5.1 | Student's t Distribution

We simulate $n \in \{100, 500, 2000\}$ observations from a Student's t distribution with $\nu \in \{1, 5, 100\}$ degrees of freedom. For each configuration, we repeat the simulation $N = 1000$ times and

compute the estimated inflection point and point of maximum convexity of the right tail using $\arg\max_{x>0} -f_n'(x)$ and $\arg\max_{x>0} f_n''(x)$, respectively. The mean squared errors (MSE) and mean absolute errors (MAE) are given in Tables 1 and 2, and as expected, they decrease with n and are larger for the sample version of PMConv_r compared with PInf_r as the former involves higher derivatives. The errors also decrease with ν since, as the distributions become closer to a Gaussian, the derivatives $f'(x)$ and $f''(x)$ fluctuate less near the origin and can be more accurately estimated by kernel density methods.

Figures 15, 16, and 17 show the estimated density functions and their first two derivatives for simulated data with parameters $n = 2000$ and $\nu = 1$. We can see in Figure 17 that the estimates oversmooth the second derivative, and to better capture the true shape of the second derivative near 0 and the PMCurv, we would need a smaller bandwidth than the one given by (8). A potential solution is offered by implementing adaptive bandwidth techniques (Politis et al. 2015) where the bandwidth depends on the location.

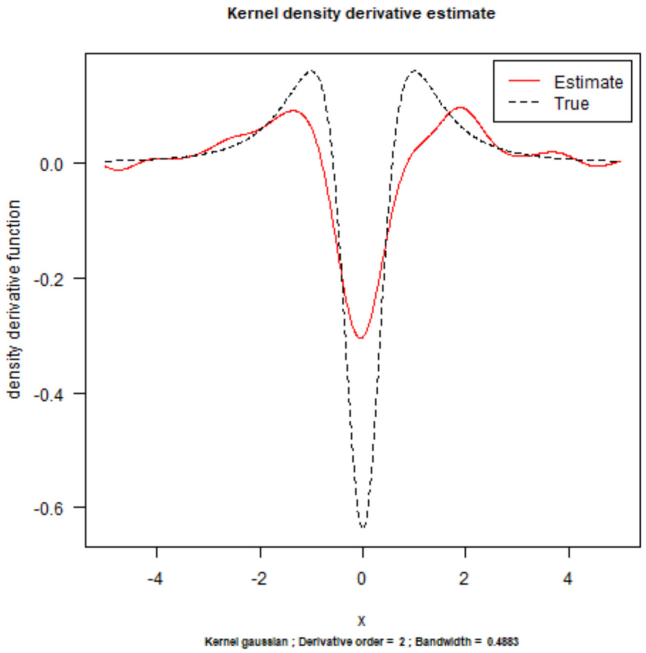


FIGURE 17 | Second derivative of the density function of a Cauchy distribution estimated from 2000 simulated observations.

5.2 | Log-Normal Distribution

As in the Student's t example, we simulate $n \in \{100, 500, 2000\}$ observations from log-normal distributions with associated mean parameter $\mu = 0$ and standard deviation $\sigma \in \{0.5, 1, 2\}$. We repeat the simulation for $N = 1000$ replicates and use the results to compute the MSE and MAE values. The results are reported in Tables 3 and 4. We observe a similar behaviour to the Student's t case, where increasing the sample size and the variability of the underlying distribution via the parameter σ, generally decreases the errors.

5.3 | Exponential Distribution

The exponential distribution presents a singularity at 0 and has no inflection point. However, we know the expression of PMCurv, as reported in Section 3.2. We simulate $n \in \{100, 500, 2000\}$ observations from exponential distributions with rate parameter $\lambda \in \{1, 2, 5\}$. We repeat the simulation for $N = 1000$ replicates and use the results to compute the MSE and MAE values. The results are reported in Table 5. The values of the errors across scenarios are comparable.

TABLE 3 | MSE and MAE for the estimated inflection points considering n simulated observations from a log-normal distribution with mean $\mu = 0$ and standard deviation σ.

	MSE			MAE		
	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 2.0$	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 2.0$
$n = 100$	0.136	0.485	0.219	0.350	0.638	0.432
$n = 500$	0.063	0.203	0.096	0.242	0.430	0.304
$n = 2000$	0.032	0.105	0.060	0.175	0.314	0.244

TABLE 4 | MSE and MAE for the estimated points of maximum convexity considering n simulated observations from a log-normal distribution with mean $\mu = 0$ and standard deviation σ.

	MSE			MAE		
	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 2.0$	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 2.0$
$n = 100$	2.197	0.814	0.316	1.482	0.800	0.413
$n = 500$	2.058	0.954	0.173	1.434	0.966	0.315
$n = 2000$	1.917	1.000	0.098	1.385	1.000	0.245

5.4 | Gaussian Distribution

In this scenario, we simulate $n \in \{100, 500, 2000\}$ observations from Gaussian distributions with mean $\mu = 0$ and standard deviation $\sigma \in \{0.5, 1, 2\}$. We repeat the simulation for $N = 1000$ replicates and use the results to compute the MSE and MAE values. The results are reported in Tables 6 and 7. As in the case of the log-normal and Student's t, we observe a decrease of the errors with increasing sample size and with decreasing the variance parameter.

6 | Conditional Excess Distribution Function

The Pickands-Balkema-de Haan Theorem (Balkema and De Haan 1974; Pickands 1975) states that, for a sufficiently large threshold, the generalized Pareto distribution (GPD) adequately approximates the conditional excess distribution function of a random variable X with cdf F. The conditional excess distribution function is defined as

$$F_u(y) = \mathbb{P}(X - u \leq y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}$$

TABLE 5 | MSE and MAE for the estimated points of maximum convexity considering n simulated observations from a Exponential distribution with rate λ.

	MSE			MAE		
	$\lambda = 1$	$\lambda = 2$	$\lambda = 5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 5$
$n = 100$	3.311	0.041	0.320	1.782	0.174	0.562
$n = 500$	2.215	0.078	0.397	1.469	0.268	0.630
$n = 2000$	1.557	0.147	0.456	1.240	0.378	0.675

where u is a given threshold. More in details, there exist two functions $a(u) > 0$ and $b(u)$ such that

$$F_u\big(a(u)y + b(u)\big) \to G_{\xi,\beta}(y), \text{as } u \to \infty$$

where $G_{\xi,\beta}(y)$ is the GPD with parameters ξ and β.

Figure 18 shows F_u evaluated for a Cauchy and a standard normal random variable, with u equal to the point of maximum curvature and the 99th percentile of the corresponding distribution. We compare F_u with the GDP approximation (Baricz 2008) and note that the point of maximum curvature as a threshold yields a better approximation.

7 | Discussion

This paper proposes using the inflection point, the point of maximum convexity (PMConv), and the point of maximum curvature (PMCurv) of probability density functions as the delimiting points between the bulk and the tail of the distribution. The inference of these delimiting points from the data is also discussed.

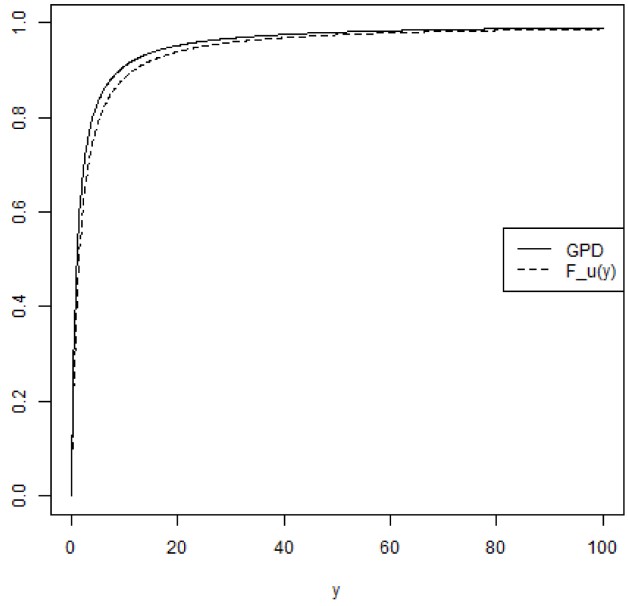
A further investigation into the relationships between these delimiting points and established measures of heavy-tailedness can provide valuable insights. Notably, it is observed that as kurtosis increases, the inflection point and point of maximum convexity tend to approach the mode and yield narrower modal regions. The previous observation holds for many common distributions-such as the Student's t-distribution, inverse-Gaussian, inverse-gamma, and various subexponential distributions like the log-normal, Weibull, and log-gamma. Conversely, the 5% and 95% quantiles generally lead to wider intervals when the kurtosis increases. The delimiting points for the right tail also tend to move closer to

TABLE 6 | MSE and MAE for the estimated inflection points considering n simulated observations from a Gaussian distribution with mean $\mu = 0$ and standard deviation σ.

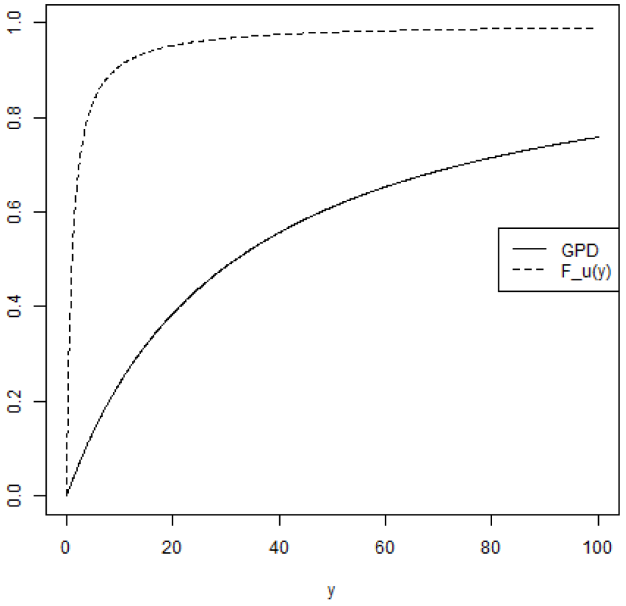
	MSE			MAE		
	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 2.0$	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 2.0$
$n = 100$	0.045	0.109	0.432	0.199	0.265	0.519
$n = 500$	0.019	0.043	0.212	0.131	0.168	0.368
$n = 2000$	0.009	0.020	0.118	0.092	0.116	0.273

TABLE 7 | MSE and MAE for the estimated points of maximum convexity considering n simulated observations from a Gaussian distribution with mean $\mu = 0$ and standard deviation σ.

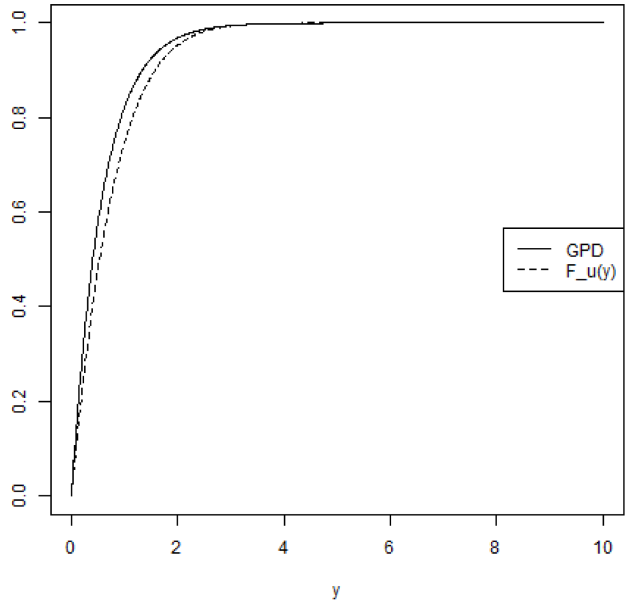
	MSE			MAE		
	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 2.0$	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 2.0$
$n = 100$	0.204	0.334	0.716	0.444	0.509	0.702
$n = 500$	0.109	0.148	0.448	0.327	0.346	0.549
$n = 2000$	0.063	0.084	0.284	0.250	0.263	0.438



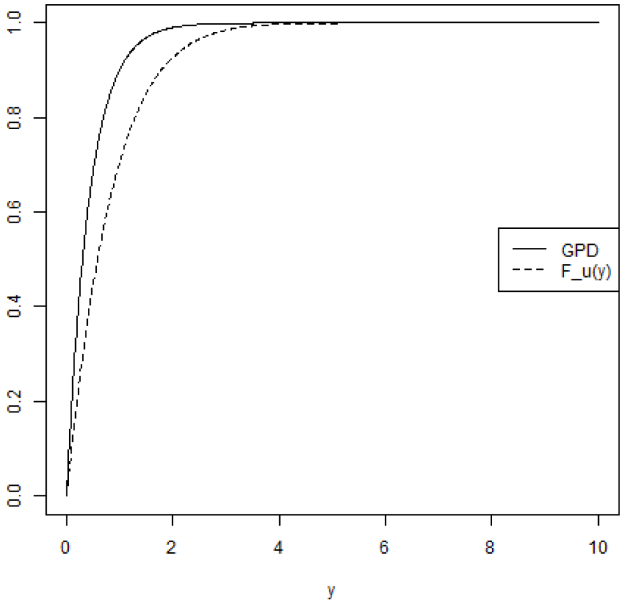
(a) $X \sim$ Cauchy, $u =$ PMCurv$_r$



(b) $X \sim$ Cauchy, $u =$ 99th percentile



(c) $X \sim$ N(0, 1), $u =$ PMCurv$_r$



(d) $X \sim$ N(0, 1), $u =$ 99th percentile

FIGURE 18 | Asymptotic tail behaviour under the Pickands-Balkema-de Haan Theorem for $X \sim$ Cauchy (top row) and $X \sim \mathrm{N}(0,1)$ (bottom row). The conditional excess distribution function F_u (dashed line) is compared with the corresponding GPD approximation. (left column) $u =$ PMCurv$_r$, (right column) $u =$ 99th percentile of X.

the mode with increasing skewness. This observation calls for further exploration into the precise connections between these statistical measures to offer a more nuanced understanding of distribution shapes and also how to extend these results to the multivariate case and multimodal distributions.

We also emphasize the importance of better understanding the properties of probability density function derivatives, a topic rarely touched upon in the literature (Sato and Yamazato 1981). Particularly, for continuous and unimodal distributions, there appears to be a regular pattern in their derivatives, which we exploit in the proof of Theorem 1 (see Appendix S1 for details).

Endnotes

[1] The tail index α characterizes tail functions with the asymptotic behaviour $\overline{F}(x) = x^{-\alpha}L(x)$ as $x \to \infty$, where $L(x)$ is a slowly varying function.

References

Asmussen, S. 2003. "Steady-State Properties of gi/g/1." In *Applied probability and Queues: Stochastic Modelling and Applied Probability*, vol. 51, 266–301. Springer.

Azzalini, A., and A. Capitanio. 2003. "Distributions Generated by Perturbation of Symmetry With Emphasis on a Multivariate Skew

T-Distribution." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 65, no. 2: 367–389.

Balkema, A. A., and L. De Haan. 1974. "Residual Life Time at Great Age." *Annals of Probability* 2, no. 5: 792–804.

Baricz, Á. 2008. "Mills' Ratio: Monotonicity Patterns and Functional Inequalities." *Journal of Mathematical Analysis and Applications* 340, no. 2: 1362–1370.

Barndorff-Nielsen, O. E. 1997. "Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling." *Scandinavian Journal of Statistics* 24, no. 1: 1–13.

Bhattacharya, P. K. 1967. "Estimation of a Probability Density Function and Its Derivatives." *Sankhyā: The Indian Journal of Statistics, Series A* 29, no. 4: 373–382.

Branco, M. D., and D. K. Dey. 2001. "A General Class of Multivariate Skew-Elliptical Distributions." *Journal of Multivariate Analysis* 79, no. 1: 99–113.

Cabral, R., D. Bolin, and H. Rue. 2023a. "Controlling the Flexibility of Non-Gaussian Processes Through Shrinkage Priors." *Bayesian Analysis* 18, no. 4: 1223–1246.

Cabral, R., D. Bolin, and H. Rue. 2023b. "Fitting Latent Non-Gaussian Models Using Variational Bayes and Laplace Approximations." *Journal of the American Statistical Association* 119, no. 548: 2983–2995.

Carvalho, C. M., N. G. Polson, and J. G. Scott. 2010. "The Horseshoe Estimator for Sparse Signals." *Biometrika* 97, no. 2: 465–480.

Chacón, J. E. and T. Duong. 2013. "Data-Driven Density Derivative Estimation, With Applications to Nonparametric Clustering and Bump Hunting." *Electronic Journal of Statistics* 7: 499–532.

Davis, R. A., K. S. Lii, and D. N. Politis. 2011. "Remarks on Some Nonparametric Estimates of a Density Function." In *Selected Works of Murray Rosenblatt*, edited by R. Davis, K. S. Lii, and D. Politis, 95–100. Springer.

do Nascimento, F. F., D. Gamerman, and H. F. Lopes. 2012. "A Semiparametric Bayesian Approach to Extreme Value Estimation." *Statistics and Computing* 22: 661–675.

Duong, T., A. Cowling, I. Koch, and M. P. Wand. 2008. "Feature Significance for Multivariate Kernel Density Estimation." *Computational Statistics & Data Analysis* 52, no. 9: 4225–4242.

Foss, S., D. Korshunov, S. Zachary, et al. 2011. *An Introduction to Heavy-Tailed and Subexponential Distributions*. Vol. 6. Springer.

A. C. Guidoum (2015), *Kernel Estimator and Bandwidth Selection for Density and Its Derivatives*, *Department of Probabilities and Statistics, University of Science and Technology, Houari Boumediene, Algeria*.

A. C. Guidoum (2020), Kernel Estimator and Bandwidth Selection for Density and Its Derivatives: The Kedd Package, *arXiv preprint arXiv:2012.06102*.

Härdle, W., J. S. Marron, and M. P. Wand. 1990. "Bandwidth Choice for Density Derivatives." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 52, no. 1: 223–232.

Huber, P. J. 2004. *Robust Statistics*. Vol. 523. John Wiley &: Sons.

Jones, M. C., J. S. Marron, and S. J. Sheather. 1996. "A Brief Survey of Bandwidth Selection for Density Estimation." *Journal of the American Statistical Association* 91, no. 433: 401–407.

Longin, F. 2016. *Extreme Events in Finance: A Handbook of Extreme Value Theory and Its Applications*. John Wiley & Sons.

MacDonald, A., C. J. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell. 2011. "A Flexible Extreme Value Mixture Model." *Computational Statistics & Data Analysis* 55, no. 6: 2137–2157.

Murphy, C., J. A. Tawn, and Z. Varty. 2024. "Automated Threshold Selection and Associated Inference Uncertainty for Univariate Extremes." *Technometrics* 67, no. 2: 215–224.

Northrop, P. J., N. Attalides, and P. Jonathan. 2017. "Cross-Validatory Extreme Value Threshold Selection and Uncertainty With Application to Ocean Storm Severity." *Journal of the Royal Statistical Society: Series C: Applied Statistics* 66, no. 1: 93–120.

Parzen, E. 1962. "On Estimation of a Probability Density Function and Mode." *Annals of Mathematical Statistics* 33, no. 3: 1065–1076.

Pickands, J., III. 1975. "Statistical Inference Using Extreme Order Statistics." *Annals of Statistics* 3, no. 1: 119–131.

Pisarenko, V., and M. Rodkin. 2010. *Heavy-Tailed Distributions in Disaster Analysis*. Vol. 30. Springer Science & Business Media.

D. N. Politis, V. A. Vasilev, P. F. Tarassenko, et al. (2015) Adaptive Estimation of Density Function Derivative.

Sato, K.-I., and M. Yamazato. 1981. "On Higher Derivatives of Distribution Functions of Class *l*." *Journal of Mathematics of Kyoto University* 21, no. 3: 575–591.

Scarrott, C., and A. MacDonald. 2012. "A Review of Extreme Value Threshold Estimation and Uncertainty Quantification." *REVSTAT-Statistical journal* 10, no. 1: 33–60.

Schuster, E. F. 1969. "Estimation of a Probability Density Function and Its Derivatives." *Annals of Mathematical Statistics* 40, no. 4: 1187–1195.

Siloko, I., O. Ikpotokin, F. Oyegue, C. Ishiekwene, and B. Afere. 2019. "A Note on Application of Kernel Derivatives in Density Estimation With the Univariate Case." *Journal of Statistics and Management Systems* 22, no. 3: 415–423.

N. N. Taleb (2020) Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications, *arXiv preprint arXiv:2001.10488*.

Tancredi, A., C. Anderson, and A. O'Hagan. 2006. "Accounting for Threshold Uncertainty in Extreme Value Estimation." *Extremes* 9: 87–106.

Teugels, J. L. 1975. "The Class of Subexponential Distributions." *Annals of Probability* 3, no. 6: 1000–1011.

Z. Varty, J. A. Tawn, P. M. Atkinson, et al. (2021) Inference for Extreme Earthquake Magnitudes Accounting for a Time-Varying Measurement Process, *arXiv preprint arXiv:2102.00884*.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.