Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Social Sciences 213

Some Investigations into the Class of Exponential Power Distributions

LUKAS ARNROTH

ACTA UNIVERSITATIS
UPSALIENSIS
2023

ISSN 1652-9030
ISBN 978-91-513-1871-4
urn:nbn:se:uu:diva-507351

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in Hörsal 2, Ekonomikum, Kyrkogårdsgatan 10, Uppsala, Thursday, 5 October 2023 at 09:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Associate Professor Matúš Maciak (Department of Probability and Mathematical Statistics Faculty of Mathematics and Physics, Charles University).

Abstract

Arnroth, L. 2023. Some Investigations into the Class of Exponential Power Distributions. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences* 213. 30 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1871-4.

In this thesis, methods are developed relating to the exponential power class of distributions.

Paper I considers Bayesian linear mixed models where the usual normality assumption is replaced by the multivariate exponential power distribution. Particular focus lies on Bayesian testing of the fixed effects.

Paper II introduces a score test for the shape parameter of the exponential power distribution. A Pitman-type local analysis is used to establish asymptotic results.

Paper III investigates quantile regression based on the skewed exponential power distribution. The bridge between standard and L_p quantiles is of particular interest.

Paper IV investigates Bayesian composite L_p-quantile regression based on the skewed exponential power distribution.

Paper V considers Bayesian composite quantile regression with a particular interest in establishing theoretical justification from a non-parametric Bayesian perspective.

Keywords: Exponential power distribution, Skewed exponential power distribution, linear mixed models, score test, Lp-quantiles, quantile regression, posterior convergence rate

Lukas Arnroth, Department of Statistics, Uppsala University, SE-75120 Uppsala, Sweden.

ISSN 1652-9030
ISBN 978-91-513-1871-4
URN urn:nbn:se:uu:diva-507351 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-507351)

Dedicated to Ragna and Björn

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I **Lukas Arnroth**, Rauf Ahmad, "A robustness evaluation of Bayesian tests for longitudinal data", Communications in Statistics–Theory and Methods, 2022, 51(24), 8754–8775

II **Lukas Arnroth**, Rauf Ahmad, "Neyman's $C(\alpha)$ test for the shape parameter of the exponential power class", Journal of Statistical Computation and Simulation, 2022, 92(9), 1823–1850

III **Lukas Arnroth**, Johan Vegelius. "Quantile regression based on the skewed exponential power distribution", Communications in Statistics – Simulation and Computation, 2023, Vol./No. ahead of print

IV Lukas Arnroth. "Bayesian composite L_p-quantile regression", unpublished manuscript

V **Lukas Arnroth**, Shaobo Jin. "Posterior rate of convergence for composite quantile regression", unpublished manuscript

Reprints were made with permission from the publishers.

Contents

1. Introduction

Imagine you have a set of bolts that you need to tighten or loosen. If you only have a single wrench, you may find that it's difficult to get the right fit for each bolt. For some, the wrench may be too big, while for others, it may be too small. In some cases, the wrench may not be able to reach the bolt at all. Now, imagine that you have a set of adjustable wrenches, each with different sizes and shapes. With this set of wrenches, you can adjust the size and shape of the wrench to fit each bolt precisely. You can tighten or loosen each bolt with ease, no matter its size or shape.

In this parable, the single wrench represents a standard distribution, which has a fixed shape and parameters. Just as the single wrench is limited in its ability to fit different bolts, a standard distribution may not be able to capture the complexity of different types of data. As distributions are an essential tool in the statistical toolkit, used to describe the probability of different outcomes or values that a random variable can take, it is crucial that we have flexible distributions. If such a distribution includes important special cases, its appeal should be all the more apparent.

The drawback of flexibility is more parameters, making inference and estimation more difficult, particularly with limited data. Additionally, if the model is too flexible, i.e., has too many parameters, it may not be possible to obtain reliable estimates of the parameters or make meaningful interpretations of the model. Finally, interpreting the results of a flexible distribution can be more challenging than with a simpler model, since the distribution may not have a clear interpretation or correspond to any well-known distributional family. In terms of wrenches and bolts, using a wrench designed to deal with any bolt imaginable would most likely be difficult on even the most simple bolt. Using an over-designed wrench (i.e., a distribution with too many parameters) to deal with a simple bolt (i.e., a dataset with limited complexity) may be inefficient and unnecessary.

The Exponential Power Distribution (EPD) is a family of probability distributions that has been widely used in statistical modeling due to its flexibility and versatility. The EPD is characterized by three parameters: location, scale and shape. The location and scale parameters are similar to those in the normal distribution and control the center and spread of the distribution. The shape parameter controls the tail-behaviour.

The appeal of the EPD class is that it includes both the Laplace and normal distributions as special cases, obtained via different values of the shape parameter. As the shape parameter is continuous, so is the deviation from Laplace and normal within the EPD class. The Laplace and normal are two fixed size wrenches, whilst the EPD is one adjustable wrench striking a simple balance between flexibility and complexity. With this in mind, it might not seem like such a exaggeration when Pogány and Nadarajah (2010) refer to the EPD as "perhaps the most applicable model in statistics".

Despite the potential benefits of using the EPD, the literature shows that both Laplace and normal distributions still dominate the field. However, a century ago, Wilson (1923) pointed out that the dominance of the normal distribution over the Laplace can be attributed to the mathematical properties of x^2 compared to $|x|$. According to Wilson, x^2 enjoys "all the laws of elementary mathematical analysis", whereas $|x|$ is not analytic and presents considerable mathematical difficulty in its manipulation. Moreover, the challenges are further exacerbated by considering the enclosing EPD class with $|x|^p$, which is not necessarily convex. Therefore, it is likely that the mathematical difficulties of the EPD, rather than its merits, are the main factors contributing to the disparity in usage between the EPD and its skewed extensions in the literature.

This thesis expands on the understanding of the EPD, emphasizing its flexibility and applicability in statistics. Paper I focuses on the robustness aspect of the EPD, which is a classical theme on the discussion of normal vs. Laplace. Paper II is theoretical, with focus on establishing asymptotically normal score tests with particular interest in the null of Laplace and normal. Papers III to V are thematically linked by the probabilistic interpretation of optimization of a particular asymmetric loss function which is equivalent to maximum likelihood estimation based on the skewed EPD (SEPD). Paper V stands out as it is the only one that focuses exclusively on a particular member of the SEPD, namely the asymmetric Laplace distribution.

2. Background

2.1 The exponential power distribution

One of the cornerstones of statistical theory is the practice of enclosing distributions within larger classes, which facilitates statistical analysis in a more general setting. Examples of such classes include the exponential family, elliptical distributions, and location and scale families. The exponential power distribution (EPD) is a lesser-known example of such a class, despite its potential usefulness in statistical analysis. A random variable from the EPD has density function

$$f_{EP}(x;\mu,\sigma,p) = \frac{K_{EP}(p)}{\sigma}\exp\left\{-\frac{1}{p}\left|\frac{x-\mu}{\sigma}\right|^{p}\right\} \tag{2.1}$$

where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter, $p > 0$ is the shape parameter and $K_{EP}(p) = \left(2p^{1/p}\Gamma(1+1/p)\right)^{-1}$ is the normalizing constant. Special cases of the EPD include the Laplace ($p = 1$) the normal ($p = 2$) and the uniform on $[\mu-\sigma, \mu+\sigma]$ ($p \to \infty$). The standard EPD with $p = 1, 2, 4$ is displayed in Figure 2.1. For a random variable X with density function (2.1), we have $E(X) = \mu$ and $E(|X-\mu|^{p})^{1/p} = \sigma$. In addition, the kurtosis of the EPD is given by $\kappa = \Gamma(1/p)\Gamma(5/p)/(\Gamma(3/p))^{2}$ illustrating that the thickness of the tails is controlled by p.

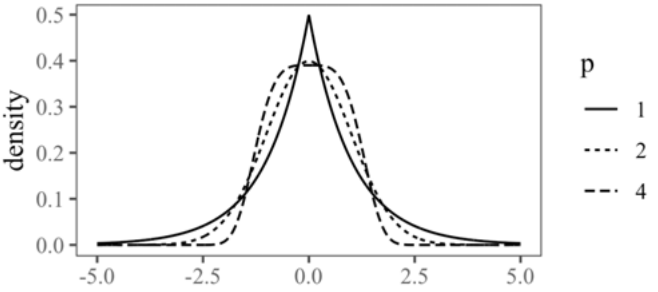


Figure 2.1. EPD with $\mu = 0$, $\sigma = 1$ for multiple values of the shape parameter

The main motivating factor for the usage of the EPD has been its special case of normality. Subbotin (1923) first introduced the EPD as a generalization of the normal distribution to model random errors. In the seminal work of Box and Tiao (1973), the EPD is considered as the source of likelihood to evaluate assumptions of normality in Bayesian inference. The normal case is contrasted with the platykurtic case, where $p > 2$, and the leptokurtic case, where $0 < p < 2$.

The EPD is also referred to as the generalized normal distribution, which is more prevalent in engineering applications, such as those described by Kwitt et al. (2010) and Lam and Goodman (2000), where Nadarajah (2005), Pogány and Nadarajah (2010), derived some important properties of the distribution.

From a theoretical standpoint, the EPD is of interest in the context of hypothesis testing, as it contains two important special cases obtained by modifying a continuous parameter. Tests based on the shape parameter, such as H_0: normal vs H_1: platykurtic or leptokurtic, are first to come to mind. However, the likelihood of a random sample with density (2.1) has to be treated with some care as the logarithm involves the L_p-distance with a known set of issues, such as points of discontinuity and non-convexity.

Daniels (1960) first considered the efficiency of the maximum likelihood estimator (MLE) of μ in (2.1), treating $\sigma = 1$ and $1/2 < p < 1$ as known. Theoretical difficulties are noted therein, and dealt with by considering differences rather than derivatives. Later, Rao (1968) utilized the EPD to investigate estimation of the mean for a continuous density at a cusp, i.e., a point of non-differentiability. In Section 8 of Le Cam and Yang (2000), a discussion based on Rao (1968) for the EPD from the perspective of local asymptotic normality is provided, where technical concepts such as differentiability in quadratic mean (QMD) is considered which served as an important starting point for paper II.

Agró (1995) was the first to consider the asymptotic distribution of MLEs of all parameters of the EPD by establishing the usual regularity conditions of the likelihood.[1] However, the technical condition of dominated third derivatives of the log-likelihood was only established for $p > 2$, effectively excluding the leptokurtic members of the EPD. Hence, in the context of hypothesis testing, we can only say that asymptotically normal Wald tests of the form H_0: normal vs H_1: platykurtic can be constructed based on the EPD.

Leptokurtic distributions, i.e., those with heavier tails than the normal distribution, are of particular importance in the context of robust statistics. The need for robust statistics arises when the assumptions of standard statistical methods are violated by the presence of outliers or other deviations from normality. Many robust methods are designed to minimize the impact of outliers on the analysis or to identify and down-weight these values when they are present. The shape parameter of the EPD can, therefore, be used as a "Winsorizing" mechanism. Barron (2019) considered M-estimation by constructing a loss function based on the log-likelihood of the EPD, where the continuous shape parameter leads to a loss function nesting many others. This perspec-

---

[1] See, e.g., Lehmann and Casella (2006) for details on the regularity conditions for asymptotically normal MLEs.

tive, giving optimization a probabilistic interpretation by viewing a objective function as a log-likelihood, has also led to what is referred to as the Bayesian bridge (Polson et al., 2014). The EPD was also studied as a source of likelihood in the Bayesian regression setting by Salazar et al. (2012); Ferreira and Salazar (2014).

We now consider the multivariate extension. A random vector $\boldsymbol{X} \in \mathbb{R}^n$ has a n-dimensional multivariate exponential power distribution (MEPD) with parameters $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$, a positive definite matrix, and $p > 0$, if its density function is given as (Gómez et al., 1998)

$$f(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},p) = k|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]^p\right\}, \qquad (2.2)$$

where

$$k = \frac{n\Gamma(n/2)}{\pi^{n/2}\Gamma\left(1+\frac{n}{2p}\right)2^{1+\frac{n}{2p}}}.$$

The function (2.2) is a special case of the elliptically contoured distributions with characteristic generator $g(t) = e^{-t^p/2}$; see, e.g. Fang et al. (1990). Some special cases are the Laplace ($p = 1/2$), normal ($p = 1$) and uniform ($p \to \infty$). A random variable $\boldsymbol{X}$ with density function (2.2) satisfies

$$E(\boldsymbol{X}) = \boldsymbol{\mu} \quad \text{and} \quad \mathrm{Var}(\boldsymbol{X}) = \frac{2^{1/p}\Gamma\left(\frac{n+2}{2p}\right)}{n\Gamma\left(\frac{n}{2p}\right)}\boldsymbol{\Sigma}.$$

The bivariate EPD is displayed for $\boldsymbol{\mu} = (0,0)^T$ and a diagonal covariance matrix for some interesting values of p in Figure 2.2.

If we consider restriction of the shape parameter to $p \in (0,1]$, then the MEPD can be conveniently reformulated in terms of a scale mixture of normal distribution, as

$$f(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},p) = \int_0^\infty N_n(\boldsymbol{y};\boldsymbol{\mu},v^2\boldsymbol{\Sigma})dH_p(v), \qquad (2.3)$$

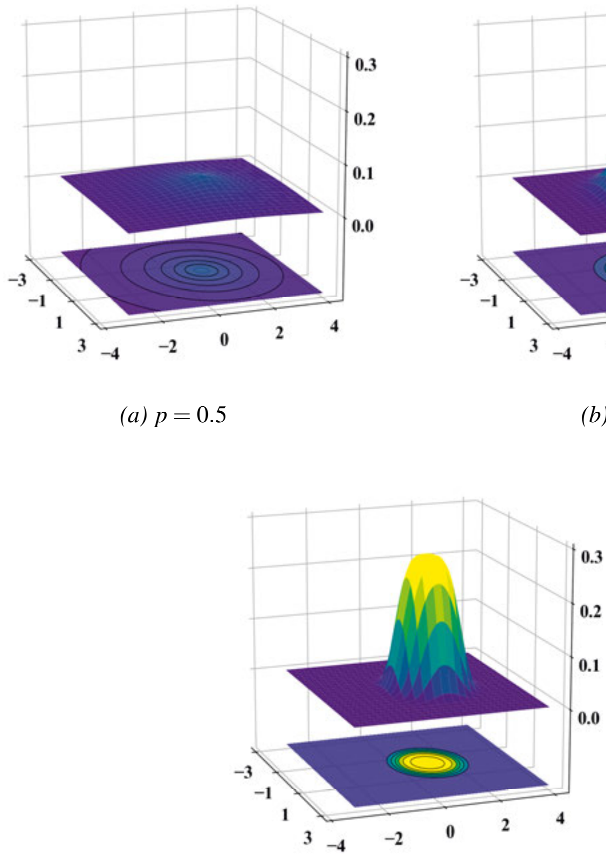
where $N_n(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes a n-variate normal distribution and H_p is a one-dimensional distribution function with density function

$$h_p(v) = \frac{2^{1+\frac{n}{2}(1-\frac{1}{p})}\Gamma\left(1+\frac{n}{2}\right)}{\Gamma\left(1+\frac{n}{2p}\right)} v^{n-3} S\left(v^{-2};p,1,\gamma_p,\delta_p\right),\ v > 0, \qquad (2.4)$$

$$\gamma_p = 2^{1-\frac{1}{p}}\cos\left(\pi\frac{p}{2}\right),\ \ \delta_p = \gamma_p \tan\left(\frac{\pi p}{2}\right).$$

Here $S\left(\cdot\ ;p,1,\gamma_p,\delta_p\right)$ in (2.4) is the density function of a stable distribution with characteristic function (Nolan, 1997)

$$\varphi(t) = \exp\left\{-\gamma_p^p|t|^p\left[1 - i\tan\left(\frac{\pi p}{2}\right)\mathrm{sign}(t)\right] + i\delta_p t\right\}.$$



(a) $p = 0.5$ *(b)* $p = 1$

(c) $p = 2$

Figure 2.2. The density function of bivariate EPD displayed for $p \in \{0.5,\ 1,\ 2\}$. Special cases of bivariate Laplace in (a) and normal in (b).

There is a wide range of applications of the MEPD, such as probabilistic deep networks (Gast and Roth, 2018), mixture modeling (Dang et al., 2015; McNicholas, 2016), image analysis (Verdoolaege and Scheunders, 2011; Luo et al., 2016; Nacereddine et al., 2019) and analysis of repeated measurements (Lindsey, 1999).

2.1.1 Application to linear mixed models

Applications in repeated measurements are of particular interest. For this, Paper I considers linear mixed models (LMM) based on the MEPD. A LMM is defined as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \quad i = 1,\ldots,m, \tag{2.5}$$

where $\boldsymbol{y}_i \in \mathbb{R}^{n_i}$ is the response vector on the ith individual, $\boldsymbol{X}_i \in \mathbb{R}^{n_i \times k}$ and $\boldsymbol{Z}_i \in \mathbb{R}^{n_i \times q}$ are the design matrices for the fixed and random effects respectively, with $\boldsymbol{\beta} \in \mathbb{R}^k$ and $\boldsymbol{b}_i \in \mathbb{R}^q$. A standard setting of LMMs assumes $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ to be independent and normally distributed as

$$\boldsymbol{b}_i \sim N_q(\boldsymbol{0}, \boldsymbol{\Psi}), \quad \boldsymbol{e}_i \sim N_{n_i}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{n_i}),$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{q \times q}$ is a symmetric positive definite matrix and $\boldsymbol{I}_{n_i}$ is the $n_i \times n_i$ identity matrix. The major advantage of using normal models for LMMs is that they allow for the simultaneous modeling of serial correlation and heterogeneity among individuals, which includes variance components or random effects. This can be achieved by appropriately structuring the covariance matrix.

If we instead assume MEPD, the above assumptions can be expressed as

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \end{bmatrix} \sim EP_{n_i+q}\left(\begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^T + \sigma^2\boldsymbol{I}_{n_i} & \boldsymbol{Z}_i\boldsymbol{\Psi} \\ \boldsymbol{\Psi}\boldsymbol{Z}_i^T & \boldsymbol{\Psi} \end{bmatrix}, p \right), \; i = 1,\ldots,m. \tag{2.6}$$

The usefulness of the scale mixture representation of the MEPD in (2.3) should now become apparent. If we restrict $p \in (0,1]$ and introduce latent independent random variables $\{v_i\}_{i=1}^m$ with density (2.4), we can express (2.6), conditional on v_i, as

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \end{bmatrix} \Bigg| \, v_i \sim N_{n_i+q}\left(\begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix}, v_i^2 \begin{bmatrix} \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^T + \sigma^2\boldsymbol{I}_{n_i} & \boldsymbol{Z}_i\boldsymbol{\Psi} \\ \boldsymbol{\Psi}\boldsymbol{Z}_i^T & \boldsymbol{\Psi} \end{bmatrix} \right), \; i = 1,\ldots,m.$$

Some similar examples, where scale mixtures of normals are utilized for LMMs, are the multivariate t-distribution (Pinheiro et al., 2001; Bai et al., 2016) and Laplace (Yavuz and Arslan, 2018).

2.2 The skewed exponential power distribution

The skewed exponential power distribution (SEPD) is an extension of the symmetric EPD, with its pdf given as (Zhu and Zinde-Walsh, 2009)

$$f_{SEP}(x;\mu,\sigma,p,\alpha) = \begin{cases} \frac{K_{EP}(p)}{\sigma} \exp\left\{ -\frac{1}{p} \left| \frac{x-\mu}{2\alpha\sigma} \right|^p \right\} & \text{if } x \leq \mu \\ \frac{K_{EP}(p)}{\sigma} \exp\left\{ -\frac{1}{p} \left| \frac{x-\mu}{2(1-\alpha)\sigma} \right|^p \right\} & \text{if } x > \mu, \end{cases} \tag{2.7}$$

where $0 < \alpha < 1$ is the asymmetry parameter. Special cases are the asymmetric Laplace distribution (ALD), skew-normal and EPD for $\alpha = 1/2$. The

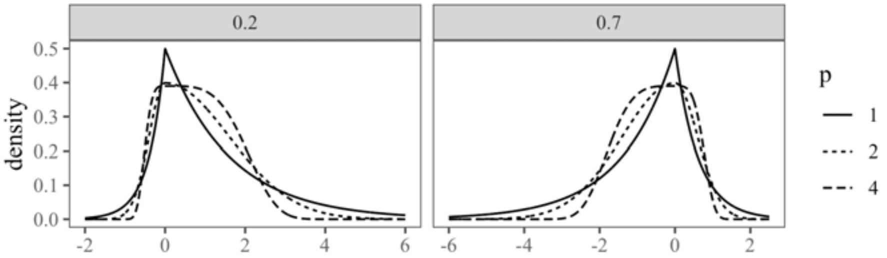


Figure 2.3. SEPD with $\mu = 0$, $\sigma = 1$ for multiple values of the shape and $\alpha = 0.2$ and 0.7

density (2.7) is displayed in Figure 2.3 for some interesting combinations of p and α.

There are multiple parametrisations of (2.7) in the literature. The two most important are outlined by Zhu and Zinde-Walsh (2009) and Komunjer (2007). They have in common that the location coincides with the α-quantile, i.e., $P(X \leq \mu) = \alpha$ for a random variable X distributed as either of the two parametrisations.

One frequent application of the SEPD is in financial modeling, where asset returns often exhibit skewness and heavy tails. The distribution can be used to model the distribution of returns and estimate the parameters of the distribution, which can then be used for risk management or portfolio optimization. Examples are Zhu and Galbraith (2011), Nadarajah et al. (2014), Chu et al. (2015).

The SEPD is used for constructing a score test for deviations from normality in terms of kurtosis and skewness by Desgagné and Lafaye de Micheaux (2018). See also Desgagné et al. (2022) for an extension of the normality test to include multiple values of the shape and skewness parameter in the null hypothesis. Their work serves, in some ways, as an extension of paper II, but can be viewed as complementary as a different theoretical approach is used.

For a random sample from the SEPD, maximum likelihood estimation of the mean is closely linked to L_p-quantiles which are based on the asymmetric L_p loss function (Chen, 1996; Daouia et al., 2019b)

$$\rho_{\alpha,p}(u) = |\alpha - I(u \leq 0)||u|^p, \quad p \geq 1. \tag{2.8}$$

L_p quantiles include both usual quantiles, with $p = 1$, and expectiles (Waltrup et al., 2015), with $p = 2$, as special cases. L_p-quantiles are used extensively in extreme value analysis, see, e.g., Usseglio-Carleve (2018); Daouia et al. (2019a,b). Much like Bayesian quantile regression is based on the asymmetric Laplace distribution (Yu and Moyeed, 2001), the SEPD can be used for Bayesian L_p-quantile regression. This thesis is not the first to note such a connection (Bernardi et al., 2018; Yang et al., 2019), but is the first to explore the relationship from the perspective of L_p-quantiles providing some important modifications to previous results.

2.2.1 Bayesian composite quantile regression

The special case of $p = 1$ in (2.7) and (2.8) is considered extensively in the context of mixtures of quantile estimators in paper V, referred to as composite quantiles. However, before any mixture of quantiles estimates is considered, we first outline the usual single quantile regression model. For the usual regression model with independent and identically distributed (iid) errors ε_i,

$$y_i = b + \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n,$$

the conditional quantiles of y are

$$Q_y(\tau|\boldsymbol{x}) = b + \boldsymbol{x}_i^T \boldsymbol{\beta} + F_\varepsilon^{-1}(\tau).$$

where $\tau \in (0,1)$ denotes the quantile of interest and $F_\varepsilon^{-1}(\cdot)$ denotes the quantile function of the errors. Quantile regression is then defined as

$$\big(\hat{b}, \hat{\boldsymbol{\beta}}(\tau)\big) = \arg\min_{b,\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(y_i - b - \boldsymbol{x}_i^T \boldsymbol{\beta}), \tag{2.9}$$

where $\rho_{\tau,1}(\cdot) =: \rho_\tau(\cdot)$. The estimator (2.9) estimates the population parameters $(b_\tau, \boldsymbol{\beta})^T$ where $b_\tau = b + F_\varepsilon^{-1}(\tau)$.

It is well known that under mild regularity conditions (Koenker, 2005)

$$\sqrt{n}\big(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\big) \xrightarrow{d} N\left(\boldsymbol{0}, \frac{\tau(1-\tau)}{f_\varepsilon(\tau)^2} \boldsymbol{C}^{-1}\right),$$

where $\boldsymbol{C} = \lim_{n\to\infty} \frac{1}{n} \sum_i \boldsymbol{x}_i \boldsymbol{x}_i^T$ and $f_\varepsilon(\cdot)$ is the density function of the errors. Hence, the relative efficiency of the quantile regression estimator with respect to the ordinary least square (OLS) estimator can be arbitrarily small depending on the quantile and distribution of the errors.

To alleviate the issue of asymptotic variance of quantile regression, Zou and Yuan (2008) proposed to simultaneously consider multiple quantile regression models. Later, Zhao and Xiao (2014) considered optimal weights for the loss functions, and we follow their definition of composite quantile regression (CQR)

$$\operatorname*{argmin}_{b_{\tau_1},\ldots,b_{\tau_K},\boldsymbol{\beta}} \sum_{i=1}^{n} \left\{ \sum_{k=1}^{K} w_k \rho_{\tau_k,1}(y_i - b_{\tau_k} - \boldsymbol{x}_i^T \boldsymbol{\beta}) \right\}, \tag{2.10}$$

where $0 < \tau_1 < \ldots < \tau_K < 1$ and $w_k > 0$, $k = 1, \ldots, K$ and $\sum_{j=1}^{K} w_j = 1$. For the Bayesian setting, Huang and Chen (2015) considered a likelihood formulated by the mixture of asymmetric Laplace distributions (ALD)

$$\prod_{i=1}^{n}\left[\sum_{k=1}^{K} w_k \frac{\tau_k(1-\tau_k)}{\sigma}\exp\left\{-\frac{1}{\sigma}\rho_{\tau_k}(y_i - b_{\tau_k} - \boldsymbol{x}_i^T\boldsymbol{\beta})\right\}\right]. \tag{2.11}$$

While paper IV deals with an extension of Bayesian CQR, paper V focuses on a theoretical evaluation of Bayesian CQR, since prior evaluations were only empirical. However, by establishing theoretical results for mixtures of the form in (2.11), many results needed for considering more general mixtures of ALD are derived. Contraction rates are a common topic for the study of the asymptotics of large and infinite dimensional mixture models. The contraction rate of a posterior distribution measures the speed at which the posterior distribution contracts around the true distribution as more data becomes available and is viewed as elaborating on its consistency. Contraction rates are of particular interest because they imply the frequentist risk of posterior estimates, thus bridging between Bayesian and frequentist estimation (Ghosal et al., 2000). Furthermore, the rate is enlightening about the way the prior distribution acts when constructing priors for high or infinite-dimensional problems (Rousseau, 2016).

3. Research aims

The primary goal of this thesis is to broaden our understanding and expand the areas of application for the (S)EPD through the development of new methods and theoretical results.

Paper I explores Bayesian LMMs under the assumption that the errors are MEPD, as opposed to the normality assumption. This work falls under the robustness branch of the EPD umbrella and focuses on hypothesis testing of the fixed-effects coefficients. It is related to the EPD-based frequentist approach for repeated measurements by Lindsey (1999).

Paper II examines score tests of the shape parameter of the EPD in both the univariate and bivariate case, in the setting of local asymptotic normality.

Papers III and IV focus on the SEPD and its relationship to quantiles. Paper III establishes a way to estimate quantiles via the SEPD by utilizing the link between L_p-quantiles and standard quantiles. Paper IV focuses on L_p-quantiles and extends Bayesian composite quantile regression to Bayesian composite L_p-quantile regression.

Lastly, Paper V focuses on the asymmetric Laplace distribution (ALD), a special case of the SEPD, which is frequently used as a likelihood in Bayesian quantile regression. The primary goal of Paper V is to establish theoretical properties of Bayesian composite quantile regression.

4. Summary of papers

4.1 Paper I

In Paper I we are interested in evaluating robustness of Bayesian tests for any number of the fixed effects parameters in Model (2.5) being 0, namely,

$$H_0 : \left(\beta_{c_1}, \ldots, \beta_{c_J}\right)^T = \mathbf{0} \text{ vs. } H_1 : \left(\beta_{c_1}, \ldots, \beta_{c_J}\right)^T \neq \mathbf{0}, \tag{4.1}$$

where $\{c_i\}_{j=1}^J \subseteq \{1, \ldots, m\}$. To test (4.1) we associate with each hypothesis models

$$\mathscr{M}_0 : y \sim f_0(y|\theta_0),\ \theta_0 \in \Theta_0 \quad \text{and} \quad \mathscr{M}_1 : y \sim f_1(y|\theta_1),\ \theta_1 \in \Theta_1$$

with corresponding prior distributions $\pi_0(\theta_0)$ and $\pi_1(\theta_1)$.

Standard Bayesian procedure is to compute the marginal likelihoods with model choice subsequently carried out using Bayes factors. Rather than computing computationally intensive marginal likelihoods, we consider testing via a two-component mixture framework recently proposed by Kamary (2016)

$$\omega f_0(y|\theta_0) + (1-\omega) f_1(y|\theta_1),\ \omega \in [0,1], \tag{4.2}$$

where θ_i are the parameters under H_i with corresponding density f_i. Posterior inference is then constructed based on the posterior distribution of ω.

In a simulation study, we treat the shape parameter as a hyper-parameter and consider the robustness of hypothesis testing of the fixed effect coefficients under varying degrees of outlier contamination. We find that the posterior distribution of ω is more robust for lower values of the shape parameter, in terms of concentration around the hypothesis of the data generating process.

The shape parameter is included in the set of unknown parameters in two applications to empirical data. In both cases, the proposed testing procedure is compared to tests based on the Hotelling's T^2 statistic and model selection via Akaike's information criterion based on frequentist estimation.

4.2 Paper II

In Paper II, we propose a Neyman's $C(\alpha)$ score test for the shape parameter of the univariate and bivariate EPD. Previous work on establishing asymptotic normality of the (S)EPD parameters show that classic regularity conditions can only be established for a subset of the parameter space, so that non-standard methods need to be considered (Agró, 1995; Zhu and Zinde-Walsh, 2009).

We use a Pitman-type local analysis, rooted in Le Cam's local asymptotic normality (LAN), which confines the alternative hypothesis to a $\sqrt{n}$-neighbourhood of the true parameter values. Fundamental to this approach is evaluating the power function at a sequence of alternatives converging to the null as sample size grows using Le Cam's third Lemma; see, e.g., van der Vaart (1998) for details. We follow Hall and Mathiason (1990) and consider tests of the form

$$H_0 : p = p_0 \quad \text{vs.} \quad H_1 : \ p = p_0 + \frac{h_p}{\sqrt{n}}$$

with unknown nuisance parameters $\lambda = (\mu, \sigma)^T$ also perturbed under the alternative as $\lambda_n = \lambda + h_\lambda / \sqrt{n}$, where h_p and h_λ are known constants. Of particular interest is the null of Laplace and normal.

For the Laplace case however, a key assumption is violated if the location μ is unknown. Hence, theoretical results for testing the null of Laplace are only established if μ is treated as known. However, we find no practical difference in the simulation study when μ is estimated.

Thus, we establish asymptotic normality with unknown nuisance parameters for lighter tails than Laplace. In a simulation study we show that the proposed tests, including that of Laplace, perform competitively relative to previous tests in the literature. In applications to empirical data, the performance of the EPD based score tests are assessed in relation to previous tests.

4.3 Paper III

In Paper III, we investigate quantile regression based on the SEPD. This idea was initially proposed by Bernardi et al. (2018) motivated by the aforementioned property of the SEPD: the location satisfies $P(X \leq \mu) = \alpha$. As in quantile regression, Bernardi et al. (2018) fix α and estimate the remaining parameters and consider the estimated regression coefficients to be quantile estimators. However, their approach should only be considered quantile regression in two special cases. Either, if the shape parameter is fixed as $p = 1$ and/or the true asymmetry parameter is known and α is set to that value.

We, instead, note that the exponential term in (2.7) contains a slight modification of the L_p-quantile loss function (2.8),

$$\rho_{p,\alpha}(u) = |\alpha - I(u \leq 0)|^{-p} |u|^p = \begin{cases} \frac{|u|^p}{\alpha^p}, & \text{if } u \leq 0 \\ \frac{|u|^p}{(1-\alpha)^p}, & \text{if } u > 0. \end{cases} \tag{4.3}$$

The L_p-quantile based on (4.3) is defined as

$$q_\alpha(p) = \arg\min_{q \in \mathbb{R}} E\rho_{p,\alpha}(Y - q). \tag{4.4}$$

If we restrict the shape parameter to $p > 1$, (4.3) is continuous and convex. The solution of (4.4) is then a 0 of $E_Y \psi_{p,\tau}(Y - q_\tau(p))$ where

$$\psi_{p,\tau}(y - q_\tau(p)) = \frac{\partial \rho_{p,\tau}(y - q_\tau(p))}{\partial q_\tau(p)} = \begin{cases} \frac{p|y-q_\tau(p)|^{p-1}}{\tau^p}, & \text{if } y - q_\tau(p) \leq 0 \\ -\frac{p|y-q_\tau(p)|^{p-1}}{(1-\tau)^p}, & \text{if } y - q_\tau(p) > 0. \end{cases}$$

Hence, by the first order condition, τ can be expressed as

$$\tau = \left[\left(\frac{E|Y - q_\tau(p)|^{p-1}}{E|Y - q_\tau(p)|^{p-1} I(Y \leq q_\tau(p))} - 1 \right)^{\frac{1}{p}} + 1 \right]^{-1}. \tag{4.5}$$

Denote the result of (4.5) if $q_\tau(p)$ is replaced with $q_\tau(1)$ by $\tilde{\tau}$. We will then have $q_{\tilde{\tau}} = q_\tau(1)$; see Daouia et al. (2019b) for details.

Bayesian and frequentist estimation procedures are outlined where this idea of the link between L_p quantiles and standard quantiles is utilized. In a simulation study, we show that the proposed method greatly outperforms that of Bernardi et al. (2018) in terms of estimating empirical quantiles. The proposed methods perform worse for estimating quantiles as compared to standard quantile regression, which is not surprising. However, the proposed methods are shown to perform better than standard quantile methods in many settings in terms of mean square error.

The same tendencies are shown in application to empirical data, where we find that the proposed method gives estimates with lower standard error but estimates quantiles worse than standard methods.

4.4 Paper IV

In Paper IV, we outline the extension of Bayesian composite quantile regression (BCQR) to L_p-quantiles. For a random sample $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$, the τth L_p-quantile regression estimate is

$$(\hat{b}_\tau, \hat{\boldsymbol{\beta}}) = \text{argmin}_{b_\tau, \boldsymbol{\beta}} \sum_{i=1}^{n} \rho_{\tau,p}(y_i - b_\tau - \boldsymbol{x}_i^T \boldsymbol{\beta}), \tag{4.6}$$

where $\rho_{\tau,p}(\cdot)$ is the asymmetric L_p loss function defined in (2.8). Following Zhao and Xiao (2014) and Huang and Chen (2015), Bayesian composite L_p-quantile regression (BCLQR) is defined as

$$(\hat{b}_{\tau_1}, \ldots, \hat{b}_{\tau_K}, \hat{\boldsymbol{\beta}}) = \text{argmin}_{b_{\tau_1}, \ldots, b_{\tau_K}, \boldsymbol{\beta}} \sum_{i=1}^{n} \left\{ \sum_{k=1}^{K} w_k \rho_{\tau_k,p}(y_i - b_{\tau_k} - \boldsymbol{x}_i^T \boldsymbol{\beta}) \right\}, \tag{4.7}$$

where quantiles $0 < \tau_1 < \ldots < \tau_K < 1$ and weights $w_k > 0$, where $\sum_{j=1}^{K} w_j = 1$, are defined as in (2.10).

For estimation we consider the likelihood for $\boldsymbol{y}$ with model matrix $\boldsymbol{X}$ as

$$p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{b},p,\sigma) = \prod_{i=1}^{n}\left(\sum_{k=1}^{K} w_k p(y_i|\boldsymbol{x}_i, b_{\tau_k}, \boldsymbol{\beta}, \sigma, \tau, p)\right), \tag{4.8}$$

where $p(\cdot)$ denotes a new version of the SEPD defined to contain (2.8) in the exponent

$$p(y|\mu,\sigma,\tau,p) = \frac{K_{\tau,p}}{\sigma}\exp\{-\rho_{\tau,p}(\sigma^{-1}(y-\mu))\}, \tag{4.9}$$

where $K_{\tau,p}^{-1} = \Gamma(1+1/p)(\tau^{-1/p} + (1-\tau)^{-1/p})$.

The parameter of interest is $\boldsymbol{\beta}$ which is given a Laplace prior to allow for variable selection. In a simulation study we show that BCLQR outperforms BCQR in terms of root mean square error. However, in some cases we found that the proposed method performed worse in terms of variable selection. In applications to empirical data we found that the proposed method performed better in terms of mean absolute prediction error and generally had estimates with lower standard deviation.

4.5 Paper V

In Paper V, BCQR is revisited from a theoretical perspective. Dirichlet mixtures of asymmetric Laplace distributions are treated in general from the perspective of Bayesian non-parameterics, with particular focus on the BCQR model.

The models of interest are

$$f_{F,\sigma}(y - \boldsymbol{x}^T\boldsymbol{\beta}) := \int \psi_{\sigma,\tau}(y - \theta - \boldsymbol{x}^T\boldsymbol{\beta})dF(\theta),$$

where $\psi_{\sigma,\tau}(\cdot - \mu) := p(\cdot|\mu,\sigma,\tau,1)$, as given by (4.9), is the density of the ALD and F denotes the mixing distribution. We consider the cases where F is either a continuous distribution with infinite support points or discrete with a finite number of support points.

We make the non-Bayesian assumption of a true distribution

$$f_0 = \int \psi_{\sigma_0,\tau_0}(y - \theta - \boldsymbol{x}^T\boldsymbol{\beta}_0)dF_0.$$

Focus is now on establishing a sequence ε_n, where $\varepsilon_n \to 0$ as $n \to \infty$, such that, for large enough $M > 0$

$$\Pi\{f_{F,\sigma,\beta} : d(f_{F,\sigma,\beta}, f_{F_0,\sigma_0,\beta_0}) > M\varepsilon_n|(Y_1,\boldsymbol{X}_1),\ldots,(Y_n,\boldsymbol{X}_n)\} \to 0,$$

where d is either the Hellinger or L_1 distance.

We first consider the case where all the components of the mixture share the same asymmetry parameters, i.e., $\tau_1 = \ldots = \tau_k$ in (4.7) and (4.8) with $p = 1$. For the finite and infinite mixtures, we find that $\varepsilon_n = n^{-1/2} \log n$ and $\varepsilon_n = n^{-1/4} (\log n)^{5/4}$ respectively.

For the case directly based on BCQR, we consider the setting where $0 < \tau_1 < \ldots < \tau_k$ are fixed but misspecified in relation to the data-generating process. We find that $\varepsilon_n = n^{-1/2} \log n$, as with the finite mixture with a shared and known asymmetry parameter.

5. Future research

The main research objective in this thesis has been to explore topics where the EPD offers interesting interpretations and applications, such as robustness, hypothesis testing and L_p-quantiles. Regarding EPD-based hypothesis tests, a natural research topic is to extend the arc set by Paper II and Desgagné et al. (2022) to consider asymptotic tests based on the multivariate EPD in a general setting. From the perspective of linear mixed models, the mixture framework for testing could be more extensively investigated under the class of multivariate EPD, for example, by including random effects in the set of hypotheses.

A majority of the papers concerns quantiles in some form. L_p-quantiles and their relationship to standard quantiles have been explored, but it remains to further investigate their relationship and weaken the distributional assumptions made in Paper III . Such work would make L_p-quantiles more accessible, since currently, they are mainly used in extreme value analysis when the quantile approaches its suprema.

Paper V can be extended to skewed exponential power distributions. Such a model would encompass both composite quantiles and L_p quantiles as special cases. Moreover, it would consolidate the work on EPD and Laplace mixtures (Scricciolo, 2011; Gao and van der Vaart, 2016) into a joint framework. Paper V has already established many necessary results to deal with the asymmetry parameter on the open unit interval.

6. Acknowledgements

First and foremost, I want to thank my main supervisor, Rauf Ahmad, on whose giant shoulders this thesis stands. I'm not entirely sure how you did it, but I'm certain that your supervision has been instrumental in shaping me into an independent and confident (or at least not entirely unconfident) researcher.

Next, I would like to thank Patrik Andersson, my co-supervisor, for always being there for discussions on whatever occupies my mind. I am grateful to Silvelyn Zwanzig, whom I consider an informal co-supervisor. Taking the Theoretical Statistics course under your guidance has been the best experience I've had as a student.

Thank you Johan Vegelius for being my perfect roommate. I can only hope to reflect a small proportion of your kindness, patience and tolerance. Thank you Paulina Jonéus and Ingrid Mattson. Your friendship alone has made all the efforts of these five years worth it. Alexander Engberg, thank you for the good times at the gym. Mattias Nordin, thank you for the chess games. Thank you Inger Persson for being a inspiration. To all my PhD colleagues, thank you for your support. Lastly, thank you to everyone at the Department of Statistics for making me feel at home.

Per Wirén, thank you for all the concerts and for choosing me as the godfather of Agnes. Zakris Strömqvist and Jonni Fidler Denis, my oldest statistics friends, thank you. Mother, Father, Mikaela, and Cornelia, thank you for your support.

Thank you Ragna Reinhammar for your unwavering support and steadfast belief in my abilities. Like Beatrice guiding Dante, you have with infinite kindness and determination guided me through the emotional highs and lows of this journey.

References

Agró, G. Maximum likelihood estimation for the exponential power function parameters. *Communications in Statistics – Simulation and Computation*, 24(2): 523–536, 1995.

Bai, X., Chen, K., and Yao, W. Mixture of linear mixed models using multivariate t distribution. *Journal of Statistical Computation and Simulation*, 86(4):771–787, 2016.

Barron, J. T. A general and adaptive robust loss function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4331–4339, 2019.

Bernardi, M., Bottone, M., and Petrella, L. Bayesian quantile regression using the skew exponential power distribution. *Computational statistics & data analysis*, 126:92–111, 2018.

Box, G. E. P. and Tiao, G. C. *Bayesian inference in statistical analysis*. Addison-Wesley, 1973.

Chen, Z. Conditional lp-quantiles and their application to the testing of symmetry in non-parametric regression. *Statistics & Probability Letters*, 29(2):107–115, 1996.

Chu, J., Nadarajah, S., and Chan, S. Statistical analysis of the exchange rate of bitcoin. *PloS one*, 10(7):e0133678, 2015.

Dang, U. J., Browne, R. P., and McNicholas, P. D. Mixtures of multivariate power exponential distributions. *Biometrics*, 71(4):1081–1089, 2015.

Daniels, H. E. The asymptotic efficiency of a maximum likelihood estimator. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab*, volume 1, 151–163, 1960.

Daouia, A., Gijbels, I., S., and Stupfler, G. Extremiles: A new perspective on asymmetric least squares. *Journal of the American Statistical Association*, 114 (527):1366–1381, 2019a.

Daouia, A., Girard, S., and Stupfler, G. Extreme m-quantiles as risk measures: From l1 to lp optimization. *Bernoulli*, 25(1):264–309, 2019b.

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., and Partchev, I. The estimation of item response models with the lmer function from the lme4 package in r. *Journal of Statistical Software*, 39:1–28, 2011.

Desgagné, A. and Lafaye de Micheaux, P. A powerful and interpretable alternative to the jarque–bera test of normality based on 2nd-power skewness and kurtosis, using the rao's score test on the apd family. *Journal of Applied Statistics*, 45(13): 2307–2327, 2018.

Desgagné, A., Lafaye de Micheaux, P., and Ouimet, F. Goodness-of-fit tests for laplace, gaussian and exponential power distributions based on λ-th power skewness and kurtosis. *Statistics*, 57(1):1–29, 2022.

Fang, K., Ng, K. W., and Kotz, S. *Symmetric multivariate and related distributions*. Chapman and Hall, London, 1990.

Fernandez, C., Osiewalski, J., and Steel, M. F. J. Modeling and inference with υ-spherical distributions. *Journal of the American Statistical Association*, 90 (432):1331–1340, 1995.

Ferreira, M. A. R. and Salazar, E. Bayesian reference analysis for exponential power regression models. *Journal of Statistical Distributions and Applications*, 1(1): 1–20, 2014.

Gao, F. and van der Vaart, A. Posterior contraction rates for deconvolution of dirichlet-laplace mixtures. *Electronic Journal of Statistics*, 10(1):608–627, 2016.

Gast, J. and Roth, S. Lightweight probabilistic deep networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3369–3378, 2018.

Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. Convergence rates of posterior distributions. *Annals of Statistics*, 28(2):500–531, 2000.

Gómez, E., Gomez-Viilegas, M. A., and Marín, J. M. A multivariate generalization of the power exponential family of distributions. *Communications in Statistics – Theory and methods*, 27(3):589–600, 1998.

Griffin, M. and Hoff, P. D. Testing sparsity-inducing penalties. *Journal of Computational and Graphical Statistics*, 29(1):128–139, 2020.

Hall, W. and Mathiason, D. On large-sample estimation and testing in parametric models. *International Statistical Review*, 58(1):77–97, 1990.

Huang, H. and Chen, Z. Bayesian composite quantile regression. *Journal of Statistical Computation and Simulation*, 85(18):3744–3754, 2015.

Kamary, K. *Non-informative priors and modelization via mixture models*. PhD thesis, PSL Research University, 2016.

Koenker, R. *Quantile regression*. Cambridge university press, 2005.

Komunjer, I. Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics*, 22(5):891–921, 2007.

Kwitt, R., Meerwald, P., and Uhl, A. Lightweight detection of additive watermarking in the dwt-domain. *IEEE Transactions on Image Processing*, 20(2):474–484, 2010.

Lam, E. Y. and Goodman, J. W. A mathematical analysis of the dct coefficient distributions for images. *IEEE Transactions on Image Processing*, 9(10): 1661–1666, 2000.

Le Cam, L. and Yang, G. L. *Asymptotics in statistics: some basic concepts*. Springer Science & Business Media, 2000.

Lehmann, E. L. and Casella, G. *Theory of point estimation*. Springer Science & Business Media, 2006.

Lindsey, J. K. Multivariate elliptically contoured distributions for repeated measurements. *Biometrics*, 55(4):1277–1280, 1999.

Luo, L., Yang, J., Qian, J., Tai, Y., and Lu, G. Robust image regression based on the extended matrix variate power exponential distribution of dependent noise. *IEEE Transactions on Neural Networks and Learning Systems*, 28(9):2168–2182, 2016.

McNicholas, P. D. Model-based clustering. *Journal of Classification*, 33:331–373, 2016.

Nacereddine, N., Goumeidane, A. B., and Ziou, D. Unsupervised weld defect classification in radiographic images using multivariate generalized gaussian mixture model with exact computation of mean and shape parameters. *Computers in Industry*, 108:132–149, 2019.

Nadarajah, S. A generalized normal distribution. *Journal of Applied statistics*, 32(7): 685–694, 2005.

Nadarajah, S., Zhang, B., and Chan, S. Estimation methods for expected shortfall. *Quantitative Finance*, 14(2):271–291, 2014.

Nolan, J. P. Numerical calculation of stable densities and distribution functions. *Communications in Statistics. Stochastic Models*, 13(4):759–774, 1997.

Pinheiro, J. C., Liu, C., and Wu, Y. N. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, 10(2):249–276, 2001.

Pogány, T. K. and Nadarajah, S. On the characteristic function of the generalized normal distribution. *Comptes Rendus Mathématique*, 348(3-4):203–206, 2010.

Polson, N. G., Scott, J. G., and Windle, J. The bayesian bridge. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 713–733, 2014.

Rao, B. L. S. P. Estimation of the location of the cusp of a continuous density. *The Annals of Mathematical Statistics*, 76–87, 1968.

Rousseau, J. On the frequentist properties of bayesian nonparametric methods. *Annual Review of Statistics and Its Application*, 3:211–231, 2016.

Salazar, E., Ferreira, M. A. R., and Migon, S., Helio. Objective bayesian analysis for exponential power regression models. *Sankhya B*, 74:107–125, 2012.

Scricciolo, C. Posterior rates of convergence for dirichlet mixtures of exponential power densities. *Electronic Journal of Statistics*, 5:270–308, 2011.

Subbotin, M. T. On the law of frequency of error. *Matematicheskii Sbornik*, 31(2): 296–301, 1923.

Usseglio-Carleve, A. Estimation of conditional extreme risk measures from heavy-tailed elliptical random vectors. *Electronic Journal of Statistics*, 12(2): 4057–4093, 2018.

van der Vaart, A. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.

Verdoolaege, G. and Scheunders, P. Geodesics on the manifold of multivariate generalized gaussian distributions with an application to multicomponent texture discrimination. *International Journal of Computer Vision*, 95:265–286, 2011.

Waltrup, L. S., Sobotka, F., Kneib, T., and Kauermann, G. Expectile and quantile regression – david and goliath? *Statistical Modelling*, 15(5):433–456, 2015.

Wilson, E. B. First and second laws of error. *Journal of the American Statistical Association*, 18(143):841–851, 1923.

Yang, T., Gallagher, C. M., and McMahan, C. S. A robust regression methodology via m-estimation. *Communications in Statistics – Theory and Methods*, 48(5): 1092–1107, 2019.

Yavuz, F. G. and Arslan, O. Linear mixed model with laplace distribution (llmm). *Statistical Papers*, 59:271–289, 2018.

Yu, K. and Moyeed, R. A. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, 2001.

Zhao, Z. and Xiao, Z. Efficient regressions via optimally combining quantile information. *Econometric Theory*, 30(6):1272–1314, 2014.

Zhu, D. and Galbraith, J. W. Modeling and forecasting expected shortfall with the generalized asymmetric student-t and asymmetric exponential power distributions. *Journal of Empirical Finance*, 18(4):765–778, 2011.

Zhu, D. and Zinde-Walsh, V. Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, 148(1):86–99, 2009.

Zou, H. and Yuan, M. Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36(3), 2008.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 213

Editor: The Dean of the Faculty of Social Sciences

Distribution: publications.uu.se
urn:nbn:se:uu:diva-507351

ACTA UNIVERSITATIS
UPSALIENSIS
2023