JIRSS (2021)
Vol. 20, No. 01, pp 61-81
DOI:10.52547/jirss.20.1.61

Asymmetric Univariate and Bivariate Laplace and Generalized Laplace Distributions

Barry. C. Arnold [1], and Matthew. A. Arvanitis [2]

[1] Department of Statistics, University of California, Riverside, USA.
[2] USDA Forest Products Laboratory, Madison, Wisconsin, USA.

Received: 25/01/2021, Revision received: 17/03/2021, Published online: 03/04/2021

Abstract. Alternative specifications of univariate asymmetric Laplace models are described and investigated. A more general mixture model is then introduced. Bivariate extensions of these models are discussed in some detail, with particular emphasis on associated parameter estimation strategies. Multivariate versions of the models are briefly introduced.

Keywords. Asymmetric Laplace, Bivariate Laplace, Exponential Minima, Gamma Components, Generalized Asymmetric Laplace, Laplace

MSC: 62E10, 62H05.

1 Introduction

For many years, researchers have been troubled by the sensitivity of normal models to the occurrence of the occasional extreme observation. The idea of utilizing least absolute deviation instead of least squares suggested the use of Laplace distributions as viable competitors of normal models. The existence of asymmetric error configurations called

Corresponding Author: Barry. C. Arnold (barnold@ucr.edu)
Matthew. A. Arvanitis (matthew.arvanitis@usda.gov)

for the development of asymmetric models. Recently (Kozubowski and Podgórski (2000a), Kozubowski and Podgórski (2000b)), attention has been drawn to the existence of asymmetric Laplace models which, when compared with normal models, have a double advantage of reduced sensitivity to outliers together with an ability to model asymmetry. Such Laplace models are reviewed and generalized in Section 2, together with discussion of suitable parameter estimation strategies. In Section 3, we consider the problem of developing flexible families of bivariate models with asymmetric Laplace marginals, together with some natural generalizations. These models provide direct competition to the usually employed classical bivariate normal model. In two dimensions, these Laplace models and their generalized versions continue to exhibit a degree of outlier resistance and an ability to adapt to asymmetries in the data. A small simulation study is included in Section 4, and in the concluding Section, discussion of analogous higher dimensional models is presented.

2 Univariate Asymmetric Laplace Models

Kozubowski and Podgórski (2000ab) define the univariate asymmetric Laplace distribution to be one with characteristic function of the form:

$$\phi_X(t) = (1 + \sigma^2 t^2 - i\mu t)^{-1}, \tag{2.1}$$

where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. This notation is a bit suspect since it may give the impression that μ and σ are location and scale parameters, which they are not. However, it is true that, if $\mu = 0$, then the distribution is symmetric.

An alternative representation of the model is available and will be used. For it we begin with two independent exponential random variables, V_1 and V_2 with $V_i \sim exp(\lambda_i)$, $i = 1, 2$. We then define

$$X = V_1 - V_2. \tag{2.2}$$

The characteristic function of X, defined by (2.2), is readily found to be of the form

$$\phi_X(t) = [1 + \lambda_1\lambda_2 t^2 - i(\lambda_1 - \lambda_2)t]^{-1}, \tag{2.3}$$

which can be recognized as the same as the Kozubowski-Podgórski characteristic function (2.1) if we set $\mu = \lambda_1 - \lambda_2$ and $\sigma^2 = \lambda_1\lambda_2$. The parameters λ_1 and λ_2 admit obvious interpretations in this version of the model, while the interpretation of μ and σ^2 appears to be problematic.

Before addressing the issue of proposing bivariate and multivariate asymmetric Laplace models, we will digress to consider an unusual property of the symmetric

Laplace distribution. The density of the univariate symmetric Laplace distribution can be obtained by two different constructions. One construction deals with the difference of two independent identically distributed exponential variables. The other construction is a 1/2 : 1/2 mixture of an exponential distribution and a negative exponential distribution. In the first case the characteristic function is:

$$(1 - i\lambda t)^{-1}(1 + i\lambda t)^{-1},$$

while in the second case, it is

$$(1/2)(1 - i\lambda t)^{-1} + (1/2)(1 + i\lambda t)^{-1}.$$

It is readily verified that the two representations agree. Jones (2019) has identified a generalization of this result by verifying that the only models in which differences of independent exponentials have the same distribution as suitable mixtures of the variables of differing signs are the asymmetric Laplace models of the form (2.1) or (2.2).

A mixture representation of (2.1) was noted by Kozubowski and Podgórski using the μ, σ^2 parameterization. It is more simply described using the λ_1, λ_2 parameters, and is as follows.

Suppose that X has a distribution of an exponential (λ_1) variable with probability $[\lambda_1/(\lambda_1 + \lambda_2)]$, and has that of the negative of an exponential (λ_2) variable with probability $[\lambda_2/(\lambda_1 + \lambda_2)]$. In that case the random variable has the asymmetric Laplace distribution (2.1)=(2.2). In such a case we will write $X \sim AL(\lambda_1, \lambda_2)$.

A convenient representation of the mixture is

$$X = IV_1 + (1 - I)(-V_2), \tag{2.4}$$

where $V_i \sim exp(\lambda_i)$, $i = 1, 2$, and I is an independent Bernoulli random variable with $P(I = 1) = [\lambda_1/(\lambda_1 + \lambda_2)]$.

The mixture representation (2.4) immediately suggests consideration of a more general model with an additional parameter for flexibility. We thus will consider

$$Y = IV_1 + (1 - I)(-V_2), \tag{2.5}$$

where $V_i \sim exp(\lambda_i)$, $i = 1, 2$, and I is an independent Bernoulli random variable with $P(I = 1) = p$.

It is readily verified that the characteristic function of a generalized asymmetric

Laplace variable of the form (2.5) is given by

$$\phi_Y(t) = \frac{1 + it[p\lambda_2 - (1-p)\lambda_1]}{1 + \lambda_1\lambda_2 t^2 + it(\lambda_2 - \lambda_1)}. \tag{2.6}$$

From this characteristic function, or from the mixture representation (2.5) we find

$$E(Y) = p\lambda_1^{-1} - (1-p)\lambda_2^{-1},$$

and

$$var(Y) = p(2-p)\lambda_1^{-2} + (1-p^2)\lambda_2^{-2} + 2p(1-p)\lambda_1^{-1}\lambda_2^{-1}.$$

In the case in which $X \sim AL(\lambda_1, \lambda_2)$ the moments simplify to become

$$E(X) = \lambda_1^{-1} - \lambda_2^{-1},$$

and

$$var(X) = \lambda_1^{-2} + \lambda_2^{-2}.$$

A key distinction between the Kozubowski-Podgórski model and the generalized asymmetric Laplace model (2.5) is that it is only in the special case in which $p = [\lambda_2/(\lambda_1 + \lambda_2)]$ that the density is continuous at 0.

Remark 1. A more general distribution than (2.5) can be constructed in which the I_i's are dependent indicators. For this more general model we begin with a random vector (J_1, J_2, J_3, J_4) with 4 possible values $(1,0,0,0), (0,1,0,0), (0,0,1,0)$ and $(0,0,0,1)$ with associated probabilities π_1, π_2, π_3 and π_4, and then define $I_1 = max\{J_1, J_3\}$ and $I_2 = max\{J_2, J_3\}$. The random variables (X, Y) are then defined as in (2.5) using the dependent I_i's just defined. The same modification can be made to generalize (2.4).

Remark 2. Higher dimensional versions can also be envisioned. However even in two dimensions the number of parameters will be $8 + 3 = 11$ while in three dimensions there will be $26 + 7 = 33$ parameters, and $80 + 15 = 95$ parameters in the 4-dimensional model. Restriction to simplified submodels will undoubtedly be adequate in almost all applications.

2.1 Estimation for The AL Distribution

2.1.1 Maximum Likelihood

Suppose we have a sample $X_1, X_2, ..., X_n$ from an asymmetric Laplace distribution with density

$$f_X(x; \lambda_1, \lambda_2) = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}[e^{-\lambda_1 x} I(x > 0) + e^{\lambda_2 x} I(x < 0)].$$

Define $U = \sum_1^n X_i I(X_i > 0)$, $V = -\sum_1^n X_i I(X_i < 0)$ and $W = \sum_1^n I(X_i > 0)$.

We wish to estimate the parameters using maximum likelihood. The mle's are given by solving the likelihood equations to get: $\widetilde{\lambda}_1 = \frac{n}{U + \sqrt{UV}}$ and $\widetilde{\lambda}_2 = \frac{n}{V + \sqrt{UV}}$.

2.1.2 Method of Moments

Suppose we have a sample $X_1, X_2, ..., X_n$ from an asymmetric Laplace distribution represented in the form

$$X = V_1 - V_2,$$

where the V_i's are i.i.d. with $V_i \sim \textit{exponential}(\lambda_i), i = 1, 2$.

We will equate the first moment of X and the first absolute moment of X to the corresponding sample moments. Elementary computations yield

$$E(X) = \lambda_1^{-1} - \lambda_2^{-1} \text{ and } E(|X|) = \frac{\lambda_1^2 + \lambda_2^2}{\lambda_1 \lambda_2 (\lambda_1 + \lambda_2)}.$$

Denote the corresponding sample moments by

$$M_1 = (1/n) \sum_1^n X_i \text{ and } M_2 = (1/n) \sum_1^n |X_i|,$$

and set up and solve the equations $E(X) = M_1$ and $E(|X|) = M_2$ to get the following MOM estimates

$$\begin{aligned}\widetilde{\lambda}_1 &= \left[\frac{M_2 + M_1 + \sqrt{M_2^2 - M_1^2}}{2}\right]^{-1},\\ \widetilde{\lambda}_2 &= \left[\frac{M_2 - M_1 + \sqrt{M_2^2 - M_1^2}}{2}\right]^{-1}.\end{aligned} \tag{2.7}$$

2.2 Bayesian Method

Suppose we have a sample $X_1, X_2, ..., X_n$ from an asymmetric Laplace distribution with density

$$f_X(x; \lambda_1, \lambda_2) = \frac{\lambda_1\lambda_2}{\lambda_1 + \lambda_2}[e^{-\lambda_1 x}I(x > 0) + e^{\lambda_2 x}I(x < 0)].$$

Define $U = \sum_1^n X_i I(X_i > 0)$, $V = -\sum_1^n X_i I(X_i < 0)$ and $W = \sum_1^n I(X_i > 0)$ We wish to estimate the parameters from a Bayesian viewpoint. The likelihood for the sample is given by

$$L(\lambda_1, \lambda_2) = \left(\frac{\lambda_1\lambda_2}{\lambda_1 + \lambda_2}\right)^n e^{-\lambda_1 u} e^{-\lambda_2 v}.$$

If we take independent gamma priors for λ_1 and λ_2, say

$$\widetilde{\lambda}_i \sim \Gamma(\alpha_i, \delta_i), \quad i = 1, 2,$$

our posterior density will be of the form

$$f(\lambda_1, \lambda_2 | \underline{X} = \underline{x}) \propto (\lambda_1 + \lambda_2)^{-n} \lambda_1^{\alpha_1+n-1} \lambda_2^{\alpha_2+n-1} e^{-(u+\delta_1)\lambda_1} e^{-(v+\delta_2)\lambda_2}.$$

From this joint posterior density for (λ_1, λ_2) the usual Bayes estimates of λ_1 and λ_2, namely $E(\lambda_1 | \underline{X} = \underline{x})$ and $E(\lambda_2 | \underline{X} = \underline{x})$, will be obtained by numerical integration.

2.3 Estimation for the GAL Distribution

2.3.1 Maximum likelihood

Suppose instead we have a sample $X_1, X_2, ..., X_n$ from a generalized asymmetric Laplace distribution of the form

$$X = IY_1 + (1 - I)(-Y_2),$$

where $I \sim Bernoulli(p)$ and the Y_i's are independent with $Y_i \sim exp(\lambda_i)$, $i = 1, 2$. Again define $U = \sum_1^n X_i I(X_i > 0)$, $V = -\sum_1^n X_i I(X_i < 0)$, and $W = \sum_1^n I(X_i > 0)$. We wish to estimate the parameters using maximum likelihood. Here too, we can solve the likelihood equations (they are easier in this case) to obtain

$$\widetilde{p} = \frac{W}{n}, \widetilde{\lambda_1} = \frac{W}{U}, \text{ and } \widetilde{\lambda_2} = \frac{n - W}{V}. \tag{2.8}$$

2.3.2 Method of Moments

Suppose we have a sample $X_1, X_2, ..., X_n$ from a generalized asymmetric Laplace distribution of the form

$$X = IV_1 + (1 - I)(-V_2),$$

where $I \sim Bernoulli(p)$ and the V_i's are independent with $V_i \sim exp(\lambda_i)$, $i = 1, 2$. We wish to estimate the parameters using the method of moments. For convenience, define $\delta_i = \lambda_i^{-1}$, $i = 1, 2$. The first three moments are:

$$\begin{aligned} E(X) &= p\delta_1 - (1 - p)\delta_2, && (2.9) \\ E(X^2) &= 2p\delta_1^2 + 2(1 - p)\delta_2^2, && (2.10) \\ E(X^3) &= 6p\delta_1^3 - 6(1 - p)\delta_2^3. && (2.11) \end{aligned}$$

Denote the sample moments by $M_i = (1/n) \sum_{j=1}^n X_j^i$, $i = 1, 2, 3$.

The moment equations to solve are:

$$\begin{aligned} M_1 &= p\delta_1 - (1 - p)\delta_2, && (2.12) \\ M_2 &= 2p\delta_1^2 + 2(1 - p)\delta_2^2, && (2.13) \\ M_3 &= 6p\delta_1^3 - 6(1 - p)\delta_2^3. && (2.14) \end{aligned}$$

It can be shown that in order to satisfy the moment equations, $\hat{\delta}_1$ must be a root of the following polynomial

$$\begin{aligned} p(x) = {} & (24M_2 - 48M_1^2)x^5 \\ & + (144M_1^3 - 8M_3 - 48M_1M_2)x^4 \\ & + (32M_1M_3 - 96M_1^4 - 48M_1^2M_2)x^3 \\ & + (36M_1M_2^2 - 4M_2M_3 - 16M_1^2M_3 + 72M_1^3M_2 - 24M_1^2M_3)x^2 \\ & + (16M_1^3M_3 + 8M_1M_2M_3 - 36M_1^2M_2^2 - 6M_2^3)x \\ & + (6M_1M_2^3 - 4M_1^2M_2M_3), \end{aligned} \tag{2.15}$$

Eliminating complex and non-positive real roots yields all possible values for $\hat{\delta}_1$. For each, we can compute the other two estimates, if they exist, i.e,.

$$\hat{\delta}_2 = \frac{M_2 - 2M_1\hat{\delta}_1}{2(\hat{\delta}_1 - M_1)}, \tag{2.16}$$

and

$$\hat{p} = \frac{M_1 + \hat{\delta}_2}{\hat{\delta}_1 + \hat{\delta}_2}. \tag{2.17}$$

Any resulting values of $(\hat{\delta}_1, \hat{\delta}_2, p)$ which are in the parameter space are MOM estimates. Clearly, under these conditions, it may be that no estimates exist or that multiple estimates exist.

2.4 Bayesian Method

Suppose instead we have a sample $X_1, X_2, ..., X_n$ from a generalized asymmetric Laplace distribution of the form

$$X = IY_1 + (1 - I)(-Y_2),$$

where $I \sim Bernoulli(p)$ and the Y_i's are independent with $Y_i \sim exp(\lambda_i)$, $i = 1, 2$. Again define

$U = \sum_1^n X_i I(X_i > 0)$, $V = -\sum_1^n X_i I(X_i < 0)$, and $W = \sum_1^n I(X_i > 0)$.

We wish to estimate the parameters from a Bayesian viewpoint. In this case the likelihood will be

$$L(p, \lambda_1, \lambda_2) = p^w \lambda_1^w e^{-\lambda_1 u}(1 - p)^{n-w} \lambda_2^{n-w} e^{-\lambda_2 v},$$

A conjugate prior with independent marginals is available in this case. Thus we can take

$$\widetilde{p} \sim Beta(\tau_1, \tau_2), \text{ and } \lambda_i \sim \Gamma(\alpha_i, \delta_i), \ \ i = 1, 2.$$

The corresponding joint posterior density has independent marginals of the same form as in the prior.

$$\widetilde{p}|\underline{X} = \underline{x} \sim Beta(\tau_1 + w, \tau_2 + n - w),$$

$$\lambda_1|\underline{X} = \underline{x} \sim \Gamma(\alpha_1 + w, \delta_1 + u),$$

and

$$\lambda_2|\underline{X} = \underline{x} \sim \Gamma(\alpha_2 + n - w, \delta_2 + v).$$

As estimates of the parameters we can take the posterior means, i.e.,

$$\hat{p}_{(B)} = \frac{\tau_1 + w}{\tau_1 + \tau_2 + n},$$

$$\hat{\lambda}_{1(B)} = \frac{\alpha_1 + w}{\delta_1 + u},$$

and

$$\hat{\lambda}_{2(B)} = \frac{\alpha_2 + n - w}{\delta_2 + v}.$$

If the hyperparameters of the prior are negligible, these will reduce to agree with the mle's

$$\hat{p} = \frac{w}{n}, \quad \hat{\lambda}_1 = \frac{w}{u} \text{ and } \hat{\lambda}_2 = \frac{n-w}{v}.$$

3 Bivariate Models

In Arnold (Arnold (2020)) two bivariate asymmetric Laplace models are described. The first bivariate asymmetric Laplace model was introduced by Arvanitis (2018) and we refer the reader to that source for detailed discussion of the model. Construction of the model begins with the components used in developing the bivariate gamma difference model. Thus we begin with 8 independent gamma variables $U_1, U_2, ..., U_8$ with $U_j \sim \Gamma(\delta_j, 1)$, $j = 1, 2, ..., 8$.. We then define (X, Y) by

$$\begin{aligned} X &= \lambda_{11}^{-1}(U_1 + U_5 + U_7) - \lambda_{12}^{-1}(U_3 + U_6 + U_8), \\ Y &= \lambda_{21}^{-1}(U_2 + U_6 + U_7) - \lambda_{22}^{-1}(U_4 + U_5 + U_8), \end{aligned} \tag{3.1}$$

where it is assumed that the constraints, $\delta_1 + \delta_5 + \delta_7 = 1$, $\delta_3 + \delta_6 + \delta_8 = 1$, $\delta_2 + \delta_6 + \delta_7 = 1$, and $\delta_4 + \delta_5 + \delta_8 = 1$, have been imposed. This model will be called the bivariate asymmetric Laplace model of the first kind and if (X, Y) is as defined in (3.1) we will write $(X, Y) \sim BAL(1)(\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \underline{\delta})$. Since there were four constraints on the δ_j's, this is an 8 parameter model. The marginal distributions depend only on the four λ parameters, thus:

$$X \sim AL(\lambda_{11}, \lambda_{12}), \quad Y \sim AL(\lambda_{21}, \lambda_{22}). \tag{3.2}$$

Moments are obtainable from the representation (3.1):

$$E(X) = \lambda_{11}^{-1} - \lambda_{12}^{-1}, \tag{3.3}$$
$$E(Y) = \lambda_{21}^{-1} - \lambda_{22}^{-1}, \tag{3.4}$$
$$var(X) = \lambda_{11}^{-2} + \lambda_{12}^{-2}, \tag{3.5}$$
$$var(Y) = \lambda_{21}^{-2} + \lambda_{22}^{-2}, \tag{3.6}$$

and

$$cov(X, Y) = -\lambda_{11}^{-1}\lambda_{22}^{-1}\delta_5 - \lambda_{12}^{-1}\lambda_{21}^{-1}\delta_6 + \lambda_{11}^{-1}\lambda_{21}^{-1}\delta_7 + \lambda_{12}^{-1}\lambda_{22}^{-1}\delta_8. \tag{3.7}$$

It is evident from (3.7), that a full range of correlations is available in this model. A submodel with non-negative correlations can be identified by setting $\delta_5 = \delta_6 = 0$.

The second bivariate asymmetric Laplace model that we will consider will utilize the closure under minimization property of the exponential distribution. For it we again begin with 8 independent random variables, $V_1, V_2, ..., V_8$ but this time we assume that they are exponentially distributed, thus $V_j \sim exp(\lambda_j), \quad j = 1, 2, ..., 8$. We then define

$$\begin{aligned} X &= min\{V_1, V_5, V_7\} - min\{V_3, V_6, V_8\}, \\ Y &= min\{V_2, V_6, V_7\} - min\{V_4, V_5, V_8\}, \end{aligned} \tag{3.8}$$

using a construction somewhat parallel to that used in the construction of models earlier described in this paper. If (X, Y) has the structure shown in (3.8) then we will write $(X, Y) \sim BAL(II)(\underline{\lambda})$ and say that it has a bivariate asymmetric Laplace distribution of the second kind with parameter vector $\underline{\lambda}$. Note that both the first kind and the second kind bivariate asymmetric Laplace distributions have an 8 dimensional parameter space. The marginal distributions of the BAL(II) distribution are of the asymmetric Laplace form. Thus:

$$X \sim AL(\lambda_1 + \lambda_5 + \lambda_7, \lambda_3 + \lambda_6 + \lambda_8), \tag{3.9}$$
$$Y \sim AL(\lambda_2 + \lambda_6 + \lambda_7, \lambda_4 + \lambda_5 + \lambda_8), \tag{3.10}$$

The moments of the BAL(II) distribution are thus given by:

$$E(X) = [\lambda_1 + \lambda_5 + \lambda_7]^{-1} - [\lambda_3 + \lambda_6 + \lambda_8]^{-1}, \tag{3.11}$$
$$E(Y) = [\lambda_2 + \lambda_6 + \lambda_7]^{-1} - [\lambda_4 + \lambda_5 + \lambda_8]^{-1}, \tag{3.12}$$
$$var(X) = [\lambda_1 + \lambda_5 + \lambda_7]^{-2} + [\lambda_3 + \lambda_6 + \lambda_8]^{-2}, \tag{3.13}$$
$$var(Y) = [\lambda_2 + \lambda_6 + \lambda_7]^{-2} + [\lambda_4 + \lambda_5 + \lambda_8]^{-2}, \tag{3.14}$$

and

$$\begin{aligned} cov(X, Y) = {} & \xi(\lambda_1 + \lambda_5, \lambda_2 + \lambda_6, \lambda_7) - \xi(\lambda_1 + \lambda_7, \lambda_4 + \lambda_8, \lambda_5) \\ & - \xi(\lambda_3 + \lambda_8, \lambda_2 + \lambda_7, \lambda_6) + \xi(\lambda_3 + \lambda_6, \lambda_4 + \lambda_5, \lambda_8), \end{aligned} \tag{3.15}$$

in which we use the notation $\xi(\tau_1, \tau_2, \tau_3)$ to denote

$$cov(min\{W_1, W_3\}, min\{W_2, W_3\}),$$

where the W_i's are independent with $W_i \sim exp(\tau_i),\ \ i = 1, 2, 3$. The value of $\xi(\tau_1, \tau_2, \tau_3)$ is most easily approximated by simulation. In fact, one could just evaluate $cov(X, Y)$ directly by simulation using the definition (3.8).

The joint distribution of $(X, Y) \sim BAL(II)$ will include a singular component since it is clear that the event $\{X = Y\}$ has positive probability in this model (it will occur,for example, if U_7 and U_8 are the smallest of the U_j's).

Competing with these models, Kozubowski and Podgórski (2000a) consider elliptically symmetric bivariate and multivariate extensions of the basic model (2.1). In their notation, the one dimensional model has the following characteristic function.

$$\phi_X(t) = (1 + \sigma^2 t^2 - i\mu t)^{-1},$$

Analogously, a random vector $\underline{X}$ is said to have an m-dimensional K-P (Kozubowski and Podgórski) distribution if its joint characteristic function is of the form

$$\phi_{\underline{X}}(\underline{t}) = [1 + (1/2)\underline{t}'\Sigma\underline{t} - i\underline{\mu}'\underline{t}]^{-1}, \tag{3.16}$$

where $\underline{\mu} \in \mathbb{R}^m$ and Σ is a non-negative definite symmetric matrix. The corresponding density function involves a Bessel function (except if m is odd).

It is natural to seek bivariate versions of the generalized asymmetric Laplace (GAL) distribution. This can be achieved by modifying the BAL(I) or the BAL(II) by the introduction of two additional probability parameters.

The generalized version of the $BAL(I)$ model may be defined as follows

$$\begin{aligned} X &= I_1\lambda_{11}^{-1}(U_1 + U_5 + U_7) - (1 - I_1)\lambda_{12}^{-1}(U_3 + U_6 + U_8), \\ Y &= I_2\lambda_{21}^{-1}(U_2 + U_6 + U_7) - (1 - I_2)\lambda_{22}^{-1}(U_4 + U_5 + U_8), \end{aligned} \tag{3.17}$$

where it is assumed that the constraints, $\delta_1 + \delta_5 + \delta_7 = 1$, $\delta_3 + \delta_6 + \delta_8 = 1$, $\delta_2 + \delta_6 + \delta_7 = 1$, and $\delta_4 + \delta_5 + \delta_8 = 1$, have been imposed, and where the I_j's are independent with $I_j \sim Bernoulli(p_j), \quad j = 1, 2$.

Means, variances and covariance of the coordinates of this random vector are not difficult to evaluate, or could be evaluated by simulation.

The generalized version of the $BAL(II)$ model may be defined as follows

$$\begin{aligned} X &= I_1[min\{(V_1, V_5, V_7\}] - (1 - I_1)[min\{V_3, V_6, V_8\}], \\ Y &= I_2[min\{(V_2, V_6, V_7\}] - (1 - I_2)[min\{V_4, V_5, V_8)\}], \end{aligned} \tag{3.18}$$

where the λ_j's are positive parameters, and where the I_j's are independent with $I_j \sim Bernoulli(p_j), \quad j = 1, 2$.

Means, variances and covariance of the coordinates of this random vector are also not difficult to evaluate, or could be evaluated by simulation.

It should be noted that, for cases other than the independent case, the generalized version of the $BAL(II)$ is a singular distribution. Specifically, in all except the independent case, some subset of the measure-zero set $\{(x, y) : |x| = |y|$ is associated with a positive total probability.

3.1 Parameter Estimation

None of the full models described in this section are expected to be useful for practical purposes. Instead, the authors expect that they will be used as a source for smaller, more manageable, models. For all four of the bivariate models discussed, only a small collection of specific cases may be described with available explicit expressions for their distribution functions, and even these distributions are rather complex. Therefore, unconventional methods for parameter estimation must be applied. These include Approximate Bayesian Computation, Modified Maximum Likelihood, Statistical Learning Algorithms, and others. Approximate Bayesian Computation (ABC) is a particularly attractive option for two reasons. First, as with all Bayesian methods, it permits the use of prior information. Second, and arguably more important, is the fact that, with models this complex, identifiability may become an issue. By defining the set of metrics according to which the best fit is selected, the researchers may maintain a degree of control over the model's usefulness. In the following sections, selected submodels will be described, and a parameter estimation method with this principle in mind for each will be suggested. However, the method of parameter estimation for models such as these should be at the discretion of the researcher for reasons already expressed.

3.2 One-Parameter Submodels

In this section, we will take a look at a pair of simple submodels of the *BAL*(*I*) and *BAL*(*II*) with similar characteristics and compare them. To begin, consider the sub-model of the *BAL*(*I*) given by restricting its parameter space as follows:

$$\begin{aligned}
&\lambda_{11} = 4, \\
&\lambda_{12} = 3, \\
&\lambda_{21} = 1, \\
&\lambda_{22} = 4, \\
&\delta_5 \in [0,1], \\
&\delta_6 = \delta_7 = \delta_8 = 0.
\end{aligned} \tag{3.19}$$

We will call this Model M1. For the *BAL*(*II*), notice that as $\lambda_j \longrightarrow 0$, $V_j \longrightarrow \infty$. Therefore, as long as at least one of the λ's in each of the four sets; $\{\lambda_1, \lambda_5, \lambda_7\}$, $\{\lambda_3, \lambda_6, \lambda_8\}$, $\{\lambda_2, \lambda_5, \lambda_8\}$, and $\{\lambda_4, \lambda_6, \lambda_7\}$; is nonzero, we may apply a similar convention to the *BAL*(*II*) as with the *BAL*(*I*): $\lambda_j = 0$ if and only if $V_j \equiv \infty$ and is, as such, larger than $max_{i:\lambda_i>0}\{V_i\}$. This effectively eliminates V_j from the model whenever $\lambda_j = 0$. Under this convention, consider the simple sub-model of the *BAL*(*II*) by restricting its parameter space as follows:

$$\begin{aligned}
&\lambda_5 \in [0,4], \\
&\lambda_1 = \lambda_4 = 4 - \lambda_5, \\
&\lambda_2 = 1, \\
&\lambda_3 = 3, \\
&\lambda_6 = \lambda_7 = \lambda_8 = 0.
\end{aligned} \tag{3.20}$$

We will call this Model M2. Then, both the M1 and M2 have only one parameter, include the independent case when that parameter is zero, and have the same fixed marginal distributions. They therefore differ only in their dependence structure, and both can exhibit only non-positive correlations. In fact, when the parameters are nonzero, both exhibit similar tail dependencies. Contour plots of the densities for specific values of δ_5 (M1) and λ_5 (M2) leading to equal correlations are shown in Figure 1. In the independent case, the contour plots would be diamond-shaped with straight-lined edges and radii along the axes proportional to inverses of the corresponding marginal parameters. For Model M1 and Model M2, we see that there is a clear tail dependency in Quadrant II, and the distinction between the shapes of these tail dependencies for each model is subtle but clearly identifiable. This tail dependency arises in both cases since dependence between X and Y is significant only when U_5 in Model M1 or V_5 in Model M2 dominates over the other component random variables. When this happens, X will tend to be positive, and Y will tend to be negative. Pearson correlations remain fairly small (in absolute value) for these families, but for both distributions, they do have a (negative) monotonic relationship with the parameter, and including additional parameters can lead to increased correlations. Additionally, given tail dependencies are present, use of alternative measures of correlation is warranted. A comparison of the three most popular measures of correlation over the parameter spaces of M1 and M2 is given in Figure 2. It should be noted that in the case that $\delta_5 = 1$ in Model M1, and $\lambda_5 = 4$ in Model M2, the resultant distributions are identical, that is, at both the lower and upper extremes of the parameter spaces, the distributions of M1 and M2 are identical. Density plots of these two distributions are given in Figure 3. For the slightly more general case with arbitrary values of $(\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22})$, the resultant $BAL(I)$ density for this upper-bound distribution ($\delta_5 = 1$) is:

$$f_{XY}(x,y) = \left[\lambda_{12}\lambda_{21}\left(1 + \frac{\lambda_{12}}{\lambda_{11}} + \frac{\lambda_{21}}{\lambda_{22}}\right)\right]^{-1} \exp\left\{\lambda_{12}x - \lambda_{21}y\right\} \cdot \begin{cases} 1, & x < 0 \wedge y > 0, \\ \exp\left\{-\left(1 + \frac{\lambda_{12}}{\lambda_{11}} + \frac{\lambda_{21}}{\lambda_{22}}\right)\max\left\{\lambda_{11}x, -\lambda_{22}y\right\}\right\}, & \text{otherwise.} \end{cases} \tag{3.21}$$

This density for the corresponding $BAL(II)$ model can be written with a simple change of parameters. Notice that this density has three rays from the origin marking where the density is not differentiable, while the independent case has four. In essence, this family focuses its dependence structure on one specific tail dependency, also known as discordance, that which involves positive values of X and negative values of Y. The shapes of the densities in the remaining three quadrants are fairly unaffected by the parameter (though not entirely over all values in the parameter spaces). This would suggest that a means of proceeding with parameter estimation for this submodel of both M1 and M2 is a statistic which is most impacted by changes in this tail dependency, e.g. Spearman's Rho or Kendall's Tau, rather than Pearson's correlation. Both of these measures are shown in Figure 2. Spearman, in both cases, appears to be the steepest, and therefore may be the best choice for parameter estimation. The simulation results generating the correlation plots can be applied to construct a one-to-one mapping between the parameter and the statistic.

3.3 Multi-Parameter Submodels

As the one-parameter model discussed in Section 3.2 would readily suggest, freeing more of the δ's in $BAL(I)$ or λ's in $BAL(II)$ can lead to more complex models with multiple tail dependencies. Due to the intractability of the joint distributions, similar approaches, yet more complex, would need to be applied as those in Section 3.2 for parameter estimation where multiple interactions between the parameter estimators would need to be accounted for.

The generalized BAL families exhibit many oddities that may be exploited by a researcher dealing with an unusual phenomenon. Consider the submodel of the generalized $BAL(II)$ with the following parameter space:

$$\begin{aligned}
&\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_8 = 1, \\
&\lambda_6 = \lambda \in (0, \infty), \\
&\lambda_7 = \lambda^{-1}, \\
&p_1 = p \in (0, 1), \\
&p_2 = 1 - p_1.
\end{aligned} \tag{3.22}$$

We will call this Model M3. Some realizations of this family are shown in Figure 4.

A method for estimating λ and p begins with realizing that:

$$\begin{aligned}
&p = p_1 = 1 - p_2, \text{ and} \\
&\lambda = \lambda_{11} - 2 = \lambda_{12} - 2 = \frac{1}{2}\left[(\lambda_{21} \pm \sqrt{\left(\lambda_{21} - 1)^2 - 4\right)}\right(,
\end{aligned} \tag{3.23}$$

where, for convenience, we have set

$$\begin{aligned}
&\lambda_{11} = \lambda_1 + \lambda_5 + \lambda_7, \\
&\lambda_{12} = \lambda_3 + \lambda_6 + \lambda_8, \\
&\lambda_{21} = \lambda_2 + \lambda_6 + \lambda_7, \text{ and} \\
&\lambda_{22} = \lambda_4 + \lambda_5 + \lambda_8.
\end{aligned} \tag{3.24}$$

The (marginal) maximum likelihood estimators for p_1, p_2, λ_{11}, λ_{12}, λ_{21}, and λ_{22} are given by (2.8). Lastly, assume priors of $p \sim Beta(\alpha, \beta)$ and $\lambda \sim \Gamma(\tau, \eta)$, where η is a rate parameter. With this construction, we can apply ABC with the goal of minimizing:

$$S_1 = [\tilde{p}_1 - (1 - \tilde{p}_2)]^2 \text{, and } S_2 = [\max\{\mathbf{W}_\lambda\} - \min\{\mathbf{W}_\lambda\}]^2 , \tag{3.25}$$

where

$$\mathbf{W}_\lambda = \left(\tilde{\lambda}_{11} - 2, \tilde{\lambda}_{12} - 2, \frac{1}{2}\left[(\tilde{\lambda}_{21} \pm \sqrt{(\tilde{\lambda}_{21} - 1)^2 - 4} \right] \right), \tag{3.26}$$

if it exists, and $\tilde{p}_1$, $\tilde{p}_2$, $\tilde{\lambda}_{11}$, $\tilde{\lambda}_{12}$, $\tilde{\lambda}_{21}$, $\tilde{\lambda}_{22}$ are the mle's of p_1, p_2, λ_{11}, λ_{12}, λ_{21}, and λ_{22}, respectively.

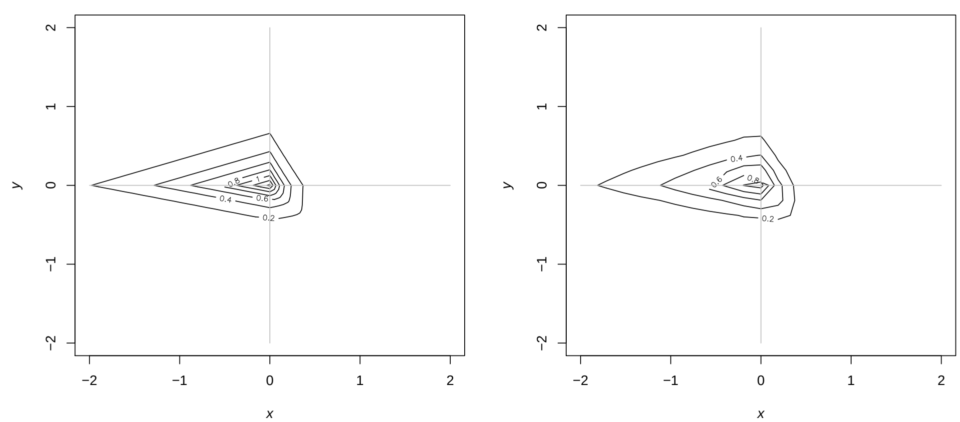

Figure 1: Contour plots of Model M1 for $\delta_5 = 0.5908$ (top) and Model M2 for $\lambda_5 = 3$ (bottom) densities. The former plot was generated using numerical integration, while the latter was computed via straight simulation.

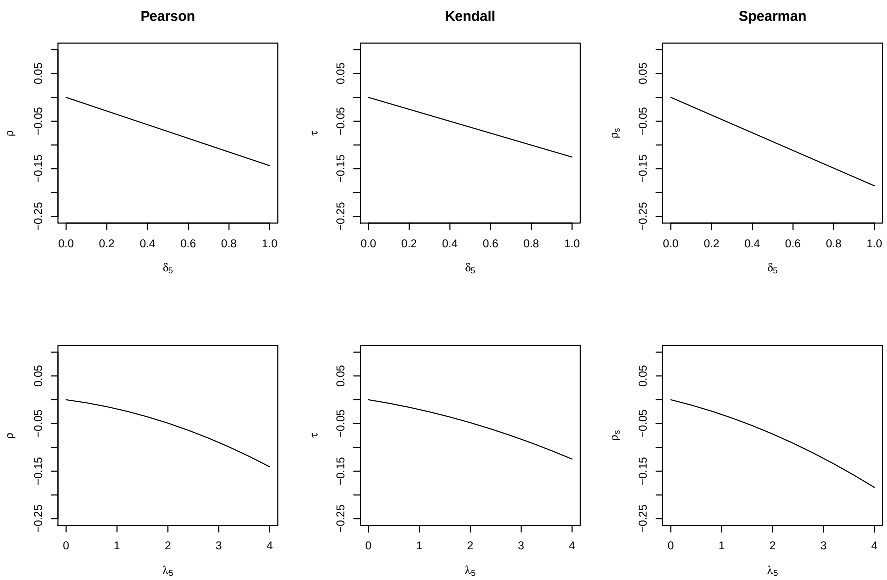

Figure 2: Correlations for M1 (top row) and M2 (bottom row) using the three methods. With exception of the Pearson for M1, which can be computed directly, these were obtained by simulation.

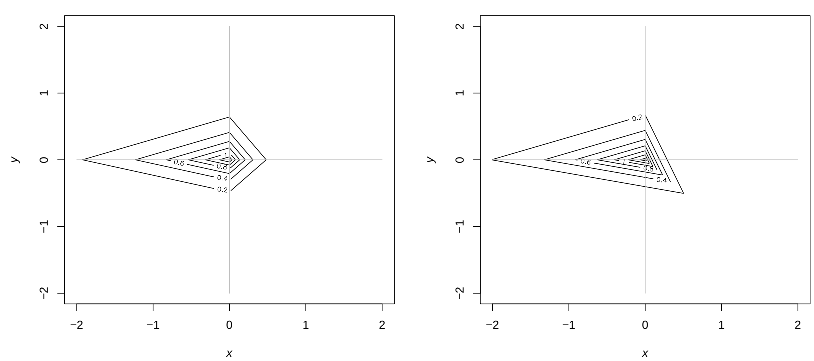

Figure 3: Densities of M1 and M2 when $\delta_5 = 0$ and $\lambda_5 = 0$, respectively (left); and when $\delta_5 = 1$ and $\lambda_5 = 4$, respectively (right).

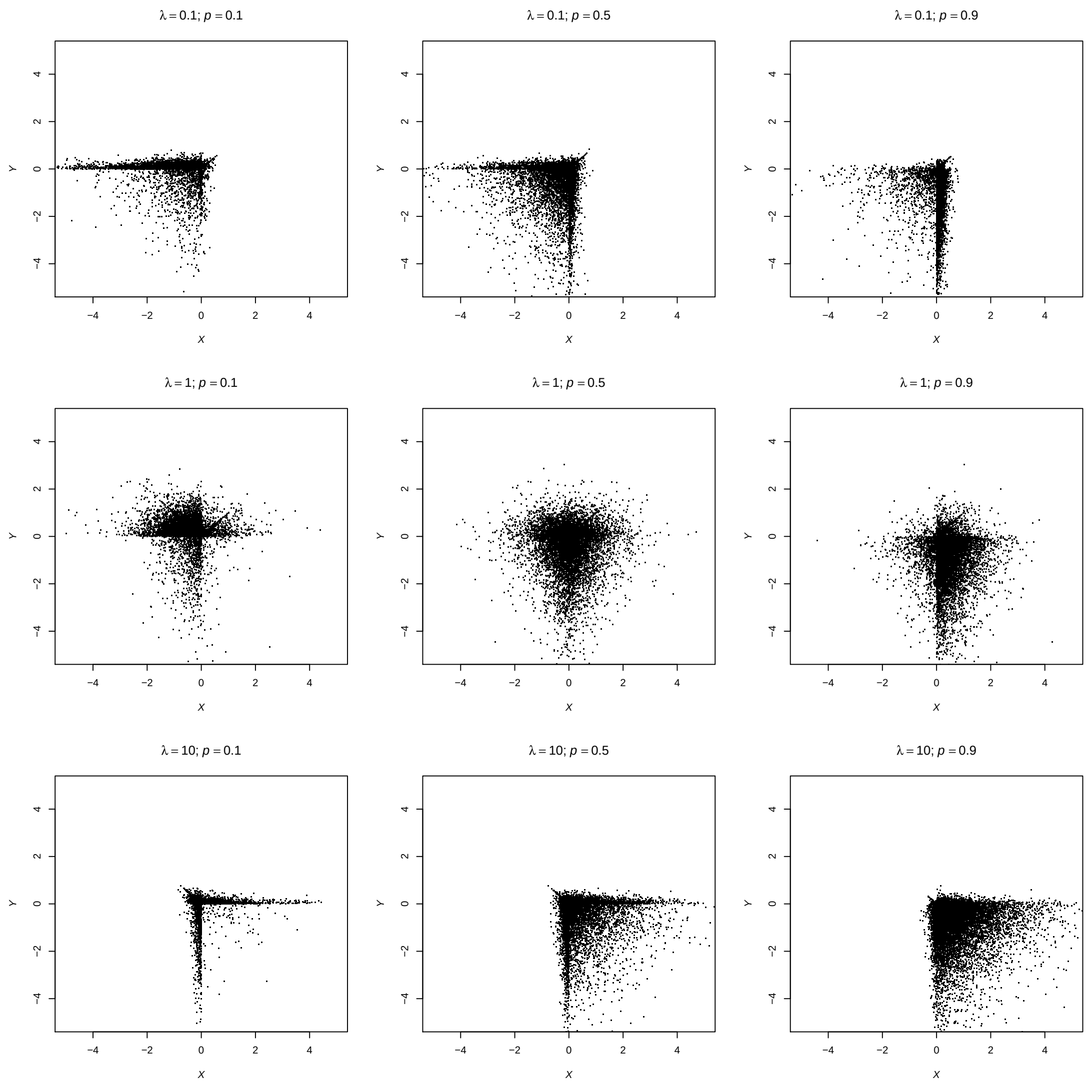

Figure 4: Realizations of Model M3 for various values of λ and p.

4 Simulations

The parameters of the $BAL(I)$ model have specific meanings. As noted previously, the λ's completely determine the marginal distributions in a convenient manner. The δ's have more subtle impacts on the distribution. To illustrate, Figure 5 shows how the δ parameters relate to specific weighted correlations.

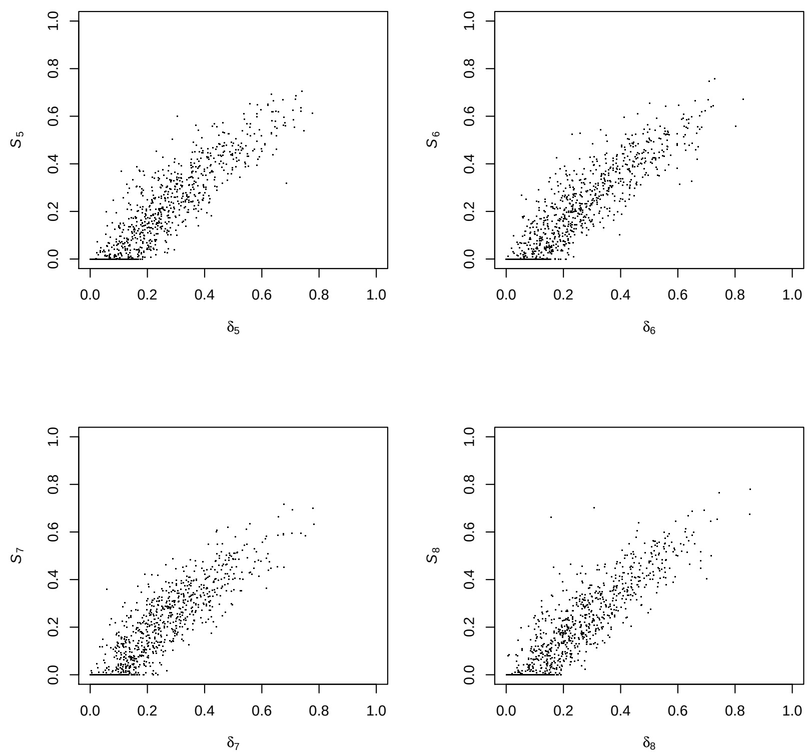

Figure 5: Relationship between δ's and weighted Pearson correlations of data in the four quadrants. For example, S_5 is computed by considering only the data in QII, and the correlation weights are simply the values of the maximum of the two coordinates (in absolute value), so that greater weight is given to points farther from the origin. S_5 and S_6 are the negatives of the weighted correlations so that all four are shown to have a positive monotonic relationship with the corresponding parameters. The correlation values were truncated at 0.

While each data point in the plots is computed from a simulated dataset of size 10,000, it should be noted that all eight parameters were free, allowing the many interactions between the parameters to create additional variation. Nonetheless, it can be seen by these plots that the parameters do have a clear monotonic relationship with this specific set of weighted correlations. The other three bivariate distributions discussed will exhibit similar (but undoubtedly more complex) relationships, but they will become more apparent when appropriate submodels are chosen for specific applications.

5 Conclusions

In this paper, two alternative specifications of univariate asymmetric Laplace models were described and investigated. Generalized versions of these were also introduced. Finally, bivariate extensions of these models were discussed in some detail, and particular emphasis was placed on associated parameter estimation strategies. We showed that these models can generate large classes of families of distributions, rich with submodels encompassing a multitude of dependence structures. The two classes of bivariate models studied both include unique benefits for researchers, particularly when unusual phenomena are studied.

Higher dimensional versions of the $BAL(I)$, $BAL(II)$, $GBAL(I)$, $GBAL(II)$ are readily described. In addition, discrete versions of the $BAL(I)$ and $BAL(II)$ models (which could be denoted by $BDAL(I)$ and $BDAL(II)$) can be defined in which geometric and negative binomial variables play the roles of the exponential and gamma variables in $BAL(I)$ and $BAL(II)$. Likewise, discrete versions of $GBAL(I)$ and $GBAL(II)$ can be described. These will be discussed in a separate report.

References

Arnold, B.C. (2020), Some bivariate and multivariate models involving independent gamma distributed components. To appear in: Contributions to Statistical Distribution Theory and Inference, a Festschrift in Honor of C. R. Rao on the Occasion of His 100th Birthday.

Arnold, B. C., Ng, and H. K. T. (2011), Flexible bivariate beta distributions. *Journal of Multivariate Analysis* **102**, 1194-1202 .

Arvanitis, M. (2018), Likelihood-Free Estimation for Some Flexible Families of Distributions (Doctoral dissertation, UC Riverside).

Jones, M. C. (2019), On a characteristic property of distributions related to the Laplace. *South African Statistical Journal* **53**, 31-34.

Kozubowski, T. J., and Podgórski, K. (2000), A Multivariate and Asymmetric Generalization of Laplace Distribution. *Computational Statistics*, **15**, 531-540.

Kozubowski, T. J., and Podgorski, K. (2000), Asymmetric Laplace distributions. *Mathematical Scientist*, **25**(1), 37-46.