South East Asian J. of Math. & Math. Sci. **ISSN : 0972-7752**
Vol.12, No.2 2016, pp. 01–18

MODIFIED SADDLEPOINT DENSITY ESTIMATES

Serge B. Provost and Susan Sheng

The University of Western Ontario,
Department of Statistical and Actuarial Sciences,
London, Ontario N6A5B7, Canada
E-mail: provost@stats.uwo.ca

Dedicated to Prof. A.M. Mathai on his 80^{th} birth anniversary

Abstract: This paper proposes a density estimation technique whereby a moment-based adjustment is applied to the saddlepoint approximation as determined from the empirical cumulant-generating function associated with a given set of observations. When two variables are involved, the product of saddlepoint density estimates of the marginal distributions is adjusted by means of a bivariate polynomial. Unlike kernel density estimates, the modified saddlepoint density estimates have simple functional representations that readily lend themselves to algebraic manipulations. Since the proposed methodology relies essentially on a determinate number of sample moments, it is particularly well suited for modeling massive data sets. As well, it should lead to improved density estimates in connection with the countless current applications arising in various fields of scientific investigation. For illustrative purposes, the density estimation approach being advocated herein is applied to two univariate and two bivariate data sets.

Keywords: Saddlepoint approximation; density estimation; moments; empirical cumulant-generating function; big data; bivariate density estimate.

2010 Mathematics Subject Classification: 62G07; 62E17; 62H10.

1. Introduction

The saddlepoint density approximation that was introduced by Daniels (1954), as well as its later refinements, may prove inaccurate when applied over the entire support of certain continuous distributions. This deficiency is illustrated in Figure 1.1 where the saddlepoint approximation of the density function of a certain mixture of normal densities previously considered by Huzurbazar (1999) is superimposed on its exact density function. This paper aims at addressing such shortcomings, mainly in the context of density estimation.

Feuerverger (1989) investigated some properties of the saddlepoint approximation in connection with the estimation of the density of a univariate sample mean. Ronchetti and Welsh (1994) extended Feuerverger's results to multivariate M-estimators. Hu *et al.* (2008) also made use of an empirical saddlepoint approximation in connection with stratified random sampling schemes. The saddlepoint approximation has been successfully applied in econometrics, biostatistics, engineering, as well as many other quantitative fields of research. An improved saddlepoint approximation was proposed by Smyth and Podlich (2002) to model the general birth process. It is still finding a variety of new statistical applications, as can been seen for example from the recently published papers of Hyrien *et al.* (2010), Kolassa and Li (2010), Demaso *et al.* (2011), Kolassa and Robinson (2011) and Marsh (2011). For an informative review on the saddlepoint methodologies, as well as their derivations and applications, the reader is referred to Reid (1988) and Butler (2010).

In light of the general density approximation results that were established in Provost (2005), it is explained in this paper that an initial saddlepoint density estimate or approximant can be improved upon by multiplying it by a moment-based polynomial adjustment. Approximation methods that are based on moments are particularly useful when the exact density of a statistical quantity cannot be obtained in closed form but its moments can be readily evaluated, as for instance is the case for numerous test statistics encountered in multivariate analysis. The univariate and bivariate cases are considered in Sections 2 and 3, respectively. Both density approximation and density estimation are discussed. The proposed methodology is then applied to several data sets in Section 4.

We conclude this introductory section by defining the saddlepoint density and distribution function approximations as defined by Daniels (1954) and Lugannani and Rice (1980), respectively. Daniels' saddlepoint approximation of the probability density function (PDF) of a continuous random variable Y at a point y belonging to its support is given by

$$g(y) = \left(2\pi\, K''(\hat{\zeta})\right)^{-1/2} \exp(K(\hat{\zeta}) - \hat{\zeta}\, y), \tag{1.1}$$

where $K(\zeta)$ is the cumulant-generating function of Y, $K''(\cdot)$ is its second derivative, and the saddlepoint $\hat{\zeta}$ is the single real solution to the equation $K'(\zeta) = y$.

The Lugannani-Rice approximation of the cumulative distribution function (CDF) of a continuous random variable Y at the point y is obtained as follows (whenever y is not equal to the expected value of Y):

$$\Pr(Y \leqslant y) \approx \Phi(\hat{w}) - \phi(\hat{w}) \left(\frac{1}{\hat{v}} - \frac{1}{\hat{w}}\right), \tag{1.2}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ respectively denote the CDF and PDF of the standard normal distribution,

$$\hat{v} = \hat{\zeta}\,(K''(\hat{\zeta}))^{1/2}$$

and

$$\hat{w} = \text{sgn}(\hat{\zeta})\,\sqrt{2(\hat{\zeta}\,y - K(\hat{\zeta}))}\,,$$

$\text{sgn}(\cdot)$ representing the sign function.

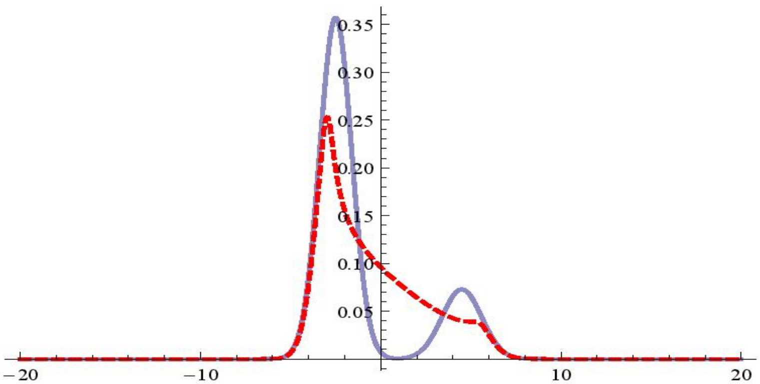

Figure 1.1: A mixture of normal PDF's and its saddlepoint approximation (dashed line).

2. Univariate Methodology

It is explained in the next two subsections that one can obtain more accurate saddlepoint density approximants or estimates by making use of certain polynomial adjustments.

2.1 Modified Saddlepoint Density Approximants

Let Y be a continuous random variable and $K(t)$ denote its cumulant-generating function. First, the Lugannani-Rice saddlepoint approximation to the distribution function of Y, as specified by Equation (1.2), is utilized in order to determine an interval whose associated probability is nearly one. For example, in the case of a semi-infinite support, one could initially choose the upper end point u in such a way that the saddlepoint distribution function evaluated at u is greater than $1 - 10^{-15}$ (or some other numerical value that is nearly equal to one). Similarly, if the lower bound of the support of the distribution were unknown, one could select ℓ, the lower end point of the initial support of the approximate distribution, to be such that the saddlepoint distribution function evaluated at ℓ is less than 10^{-15}. For an infinite support, both end points ℓ and u would have to be determined in this manner.

Saddlepoint density approximations are then evaluated from Equation (1.1) at multiple points within this initial support (for instance, fifty equidistant points) and a second order interpolating spline, $g(y)$, which is fitted to these points, serves as a continuous representation of an initial approximation to the density curve. The points of intersection of this spline with the abscissa determine the support $(l^\star, u^\star)$ of the initial approximate density function, $g^*(y)$, which is obtained by normalizing the spline on that support and referred to a the base density function. Note that such a spline can be readily obtained and manipulated algebraically, for example by utilizing the symbolic computation software *Mathematica*, which treats it as a single function.

Finally, a polynomially adjusted density approximation of the following form is assumed:

$$\psi_d(y) = g^*(y) \sum_{j=0}^{d} \xi_j \, y^j \, . \tag{2.1}$$

The polynomial coefficients ξ_j are determined by equating $\mu(h)$, the h^{th} moment of Y, which for instance can be obtained as $\frac{\partial^h e^{K(t)}}{\partial t^h}\Big|_{t=0}$, to the h^{th} moment determined from the approximate distribution specified by $\psi_d(y)$. That is,

$$\begin{aligned} \mu(h) &= \int_{l^\star}^{u^\star} y^h \, g^*(y) \sum_{j=0}^{d} \xi_j \, y^j \, \mathrm{d}y \\ &= \sum_{j=0}^{d} \xi_j \int_{l^\star}^{u^\star} y^{h+j} g^*(y) \, \mathrm{d}y \\ &\equiv \sum_{j=0}^{d} \xi_j \, m(h+j), \quad h = 0, 1, \ldots, d, \end{aligned} \tag{2.2}$$

where $m(h)$ denotes the h^{th} moment associated with $g^*(y)$. This yields a system of $d+1$ linear equations whose solution is

$$\begin{pmatrix} \xi_0 \\ \xi_1 \\ \vdots \\ \xi_d \end{pmatrix} = \begin{pmatrix} m(0) & m(1) & \cdots & m(d-1) & m(d) \\ m(1) & m(2) & \cdots & m(d) & m(d+1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ m(d) & m(d+1) & \cdots & m(2d-1) & m(2d) \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mu(1) \\ \vdots \\ \mu(d) \end{pmatrix} . \tag{2.3}$$

The integrated squared difference ($\mathcal{ISD}$) between approximations of degrees δ and $\delta+1$ is proposed as a means of selecting a suitable degree for the polynomial adjustment. More specifically, upon evaluating

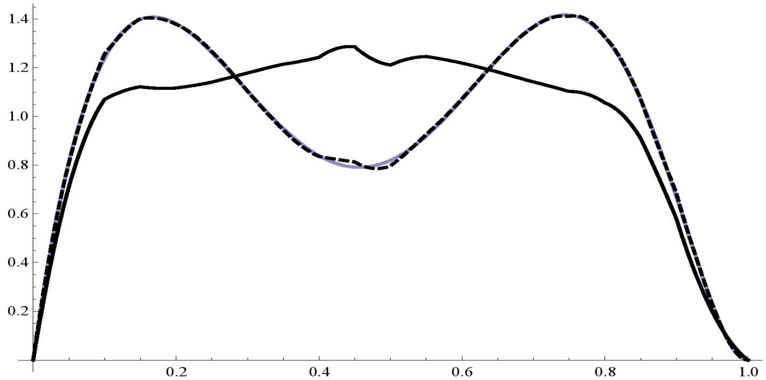

Figure 2.1: Unadjusted saddlepoint approximant (black line) and polynomially adjusted saddlepoint approximant of degree 16 (dashed line) superimposed on a given mixture of beta densities (grey line).

$$\mathcal{ISD}(\delta) = \int_{l^\star}^{u^\star} (\psi_\delta(y) - \psi_{\delta+1}(y))^2 \, \mathrm{d}y \tag{2.4}$$

for $\delta = 3, \ldots, 20$, one would select the degree $d = \delta$ for which $\mathcal{ISD}(\delta)$ is seen to converge or to reach a predetermined tolerance level. When the density function to be approximated, say $f(y)$, is known but a simpler representation is desirable, the proposed approach can be employed in conjunction with a stopping criterion that would be based on the integrated squared differences between $f(y)$ and $\psi_\delta(y)$. When the cumulant-generating function is unavailable, it can be approximated by evaluating

$$\ln\left(\sum_{j=1}^{b} \frac{t^j}{j!}\,\mu(j)\right) \tag{2.5}$$

where b should preferably be at least 20.

We close this section with an example. Consider an equal mixture of beta density functions with parameters $(2, 4)$ and $(7, 3)$, whose exact density as well as its adjusted and unadjusted approximations are plotted in Figure 2.1. In this case, a polynomial adjustment of degree 16 is seen to dramatically improve the accuracy of the saddlepoint approximant.

2.2 Adjusted Saddlepoint Density Estimates

The density approximation technique described in Section 2.1 is now adapted to the context of density estimation. Given the set of observations $y_1, y_2, \ldots, y_n$, the empirical cumulant-generating function, as defined for instance in Feuerverger

(1989), is given by

$$K^*(t) = \ln\left(\sum_{i=1}^{n} \frac{e^{t\,y_i}}{n}\right). \tag{2.6}$$

Making use of an estimate of a cumulant-generating function in lieu of an exact one enables one to obtain saddlepoint-type density estimates that can be polynomially adjusted as well. In this case, the normalized spline, $g^*(y)$, is adjusted by means of a polynomial whose coefficients are obtained from Equation (2.3) wherein $\mu(h) = \sum_{i=1}^{n} y_i^h/n$. Thus, the proposed approximation methodology can also be utilized for estimating density functions by making use of sample moments instead of exact moments. Note that unlike kernel density estimates, the resulting density estimates have simple functional forms. As a stopping rule, one can then select the degree $d = \delta$ for which $\mathcal{ISD}(\delta)$ as defined in Equation (2.4) attains a minimum value. Alternatively, the degree of the adjustment of a density estimate could be determined by re-expressing the polynomial adjustment in terms of a linear combination of the orthogonal polynomials associated with the base density and proceeding as explained in Jiang and Provost (2011).

3. Bivariate Methodology

In the bivariate case, the univariate methodology is first applied without adjustments to each component of the standardized vector, the product of the resulting density functions serving as a base density, which is then multiplied by a moment-based bivariate polynomial adjustment. On making use of the inverse of the standardizing transformation, one finally obtains the requisite density approximant or estimate, as the case may be.

3.1 Standardization and Base Density

Let $(\mu_U, \mu_V)'$ and Σ denote the mean and covariance matrix of a continuous random vector $(U, V)'$ whose support is $(a, b)\times(c, d)$. First, we apply the standardizing transformation,

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \Sigma^{-\frac{1}{2}} \begin{pmatrix} U - \mu_U \\ V - \mu_V \end{pmatrix}, \tag{3.1}$$

where $\Sigma^{-\frac{1}{2}}$ denotes the inverse of the symmetric square root of the covariance matrix Σ. Then, unadjusted saddlepoint approximants, which are represented by normalized splines, are utilized to model the respective density functions of X and Y whose product, that is, $g_X^*(x)\, g_Y^*(y)$, provides an initial bivariate density approximant that serves as base density.

Given a set of bivariate observations (u_k, v_k), $k = 1, \ldots, n$, one would replace μ_U, μ_V and Σ by their maximum likelihood estimates in Equation (3.1) in order to obtain the transformed data points (x_k, y_k), $k = 1, \ldots, n$, wherefrom the associated empirical cumulant-generating functions and sample moments can be determined for each component. Then, one would take the product of the resulting component-wise density estimates, $g_X^*(x)$ and $g_Y^*(y)$, as initial bivariate density estimate.

3.2 Bivariate Polynomial Adjustment and Inverse Transformation

First, we note that whenever the joint density function of U and V, $h_{U,V}(\cdot,\cdot)$, is known, the joint density function of X and Y can be expressed as

$$f_{X,Y}(x,y) = h_{U,V}(\beta_{11}x + \beta_{12}y + \mu_U, \beta_{21}x + \beta_{22}y + \mu_V) \times |\Sigma|^{\frac{1}{2}}, \tag{3.2}$$

where $|\Sigma|^{\frac{1}{2}}$ is the Jacobian of the inverse transformation,

$$\begin{pmatrix} U \\ V \end{pmatrix} = \Sigma^{\frac{1}{2}} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \mu_U \\ \mu_V \end{pmatrix}, \tag{3.3}$$

with $\Sigma^{\frac{1}{2}} \equiv \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}$.

The initial approximation or estimate of the joint density function of X and Y, that is, of $g_X^*(x)g_Y^*(y)$, is then adjusted by means of a bivariate polynomial of orders t and t denoted by $p_t(x,y)$ whose coefficients are such that the joint moments, as determined from $g_X^*(x)\, g_Y^*(y)\, p_t(x,y)$, coincide with those of X and Y or the joint sample moments of the transformed data, (x_k, y_k), $k = 1, \ldots, n$, in the case of a density estimate. That is, we let

$$\int_c^d \int_a^b x^i\, y^j\, g_X^*(x)\, g_Y^*(y)\, p_t(x,y)\, \mathrm{d}x\, \mathrm{d}y = \mu_{i,j} \tag{3.4}$$

for $i = 1, \ldots, t$ and $j = 1, \ldots, t$, where

$$\mu_{i,j} = E[X^i Y^j] = \int_c^d \int_a^b x^i\, y^j\, f_{X,Y}(x,y)\, \mathrm{d}x\, \mathrm{d}y,$$

$f_{X,Y}(x,y)$ being as given in Equation (3.2) or

$$\mu_{i,j} = \sum_{k=1}^{n} x_k^i\, y_k^j / n$$

when one seeks a density estimate. On letting

$$m_{h,k} = \int_c^d \int_a^b x^h\, y^k\, g_X^*(x)\, g_Y^*(y)\, \mathrm{d}x\, \mathrm{d}y$$

be the joint moments of orders h and k associated with the base density function $g_X^*(x)\,g_Y^*(y)$ and

$$p_t(x,y)=\sum_{s=0}^{t}\sum_{r=0}^{t} c_{r,s}\,x^r y^s \tag{3.5}$$

in (3.4), and then interchanging the sums and the integrals, one has

$$\sum_{s=0}^{t}\sum_{r=0}^{t} c_{r,s}\,m_{i+r,j+s}=\mu_{i,j}\,, \tag{3.6}$$

for $i=0,1,\ldots,t$ and $j=0,1,\ldots,t$, which, in matrix notation, can be expressed as

$$M\,\mathbf{c}=\boldsymbol{\mu}\,,$$

that is,

$$\begin{pmatrix} m_{0,0} & \cdots & m_{t,0} & \cdots & m_{0,t} & \cdots & m_{t,t} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ m_{t,0} & \cdots & m_{2t,0} & \cdots & m_{t,t} & \cdots & m_{2t,t} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ m_{0,t} & \cdots & m_{t,t} & \cdots & m_{0,2t} & \cdots & m_{t,2t} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ m_{t,t} & \cdots & m_{2t,t} & \cdots & m_{t,t+1} & \cdots & m_{2t,2t} \end{pmatrix} \begin{pmatrix} c_{0,0} \\ \vdots \\ c_{t,0} \\ \vdots \\ \vdots \\ c_{0,t} \\ \vdots \\ c_{t,t} \end{pmatrix} = \begin{pmatrix} \mu_{0,0} \\ \vdots \\ \mu_{t,0} \\ \vdots \\ \vdots \\ \mu_{0,t} \\ \vdots \\ \mu_{t,t} \end{pmatrix}, \tag{3.7}$$

where $\mathbf{c}$ and $\boldsymbol{\mu}$ are $(t+1)^2$-dimensional vectors whose components appear in the same order and M is a $(t+1)^2\times(t+1)^2$ matrix whose rows are reflecting the order of the components of the vector $\boldsymbol{\mu}$ in accordance with Equation (3.6). The $c_{i,j}$'s are then determined by solving the linear system specified by Equation (3.7) or, equivalently, by evaluating $M^{-1}\boldsymbol{\mu}$. Accordingly, a joint density approximant or estimate of the following form is assumed for X and Y:

$$\begin{aligned} f_t(x,y) &= g_X^*(x)\,g_Y^*(y)\,(\mathbf{c}\cdot\mathbf{z}(x,y)) \\ &\equiv g_X^*(x)\,g_Y^*(y)\,p_t(x,y) \end{aligned} \tag{3.8}$$

where $\mathbf{z}(x,y)=(1,x,\ldots,x^t,\ldots,y^t,\ldots,x^t\,y^t)'$ and $\mathbf{c}=(c_{0,0},\ c_{1,0},\ \ldots,c_{t,0},\ \ldots,\ c_{0,t},\ \ldots,c_{t,t})'$. Then, on applying the inverse transformation, which is given by

Equation (3.3), one obtains the resulting joint density approximant or estimate of the distribution of the original variables U and V as follows:

$$h_t(u,v) = f_t(x,y)(\alpha_{11}(u-\mu_U)+\alpha_{12}(v-\mu_V), \alpha_{21}(u-\mu_U)+\alpha_{22}(v-\mu_V))/|\Sigma|^{\frac{1}{2}} \tag{3.9}$$

where $\Sigma^{-\frac{1}{2}} \equiv \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}$.

The integrated squared difference between $h_t(u,v)$ and $h_{t+1}(u,v)$ (or that between $h_t(u,v)$ and $h_{U,V}(u,v)$ when the latter is known) can be utilized to select a suitable value of t, paralleling the univariate criteria. It should be pointed out that the proposed bivariate methodology can be readily extended to the multivariate setting.

4. Applications

4.1 Univariate Applications

The Buffalo Snowfall Data Set

Consider the data set consisting of the annual snowfall accumulations in Buffalo from 1910 to 1973 (available from the R package *gss*). Its distribution is modeled by making use of the density estimation methodology described in Section 2.2. The support of the approximate distribution was determined to be the interval $(-22.44, 173.84)$. Then, saddlepoint density estimates were evaluated at multiple points within that interval by making use of Equation (1.1) in conjunction with the empirical cumulant-generating function specified by Equation (2.5). The resulting normalized spline was adjusted by means of a 10^{th} degree polynomial whose coefficients were evaluated from Equation (2.3). The adjusted saddlepoint density estimate so obtained as well as a kernel density estimate (whose bandwidth was determined by Silverman's rule of thumb) are superimposed on a histogram of the data in Figure 4.1. The corresponding cumulative distribution functions are plotted in Figure 4.2 along with the empirical distribution function.

Flood Peaks

The 'flood data', previously considered by Yue (2001), consists of 77 flood peaks and volumes, as observed in the Madawaska basin. Only the peaks are considered in this application. In this case, the normalized saddlepoint spline, $g^*(y)$, was adjusted by making use of an 8^{th} degree polynomial. The resulting density estimate as well as a kernel density estimate are plotted in Figure 4.3 along with a histogram of the data. The corresponding estimated distribution functions are superimposed on the empirical distribution function in Figure 4.4.

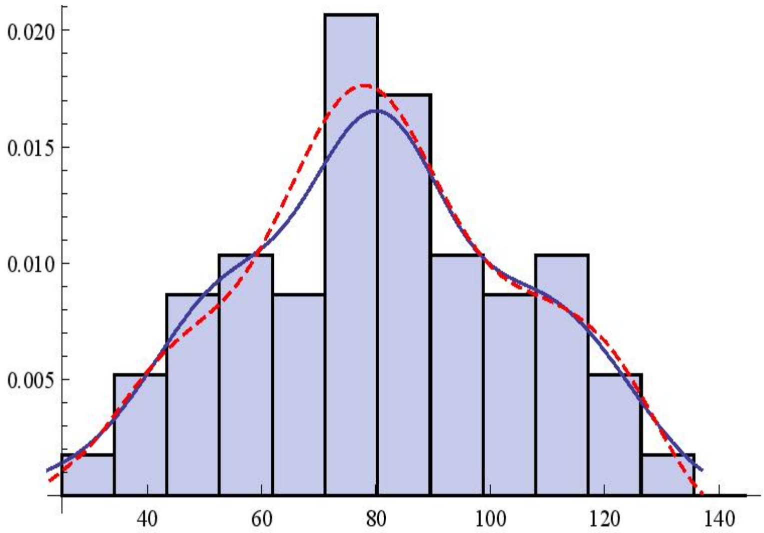

Figure 4.1: Adjusted saddlepoint density estimate (dashed line) and kernel density estimate (solid line) superimposed on a histogram of the Buffalo snowfall data.

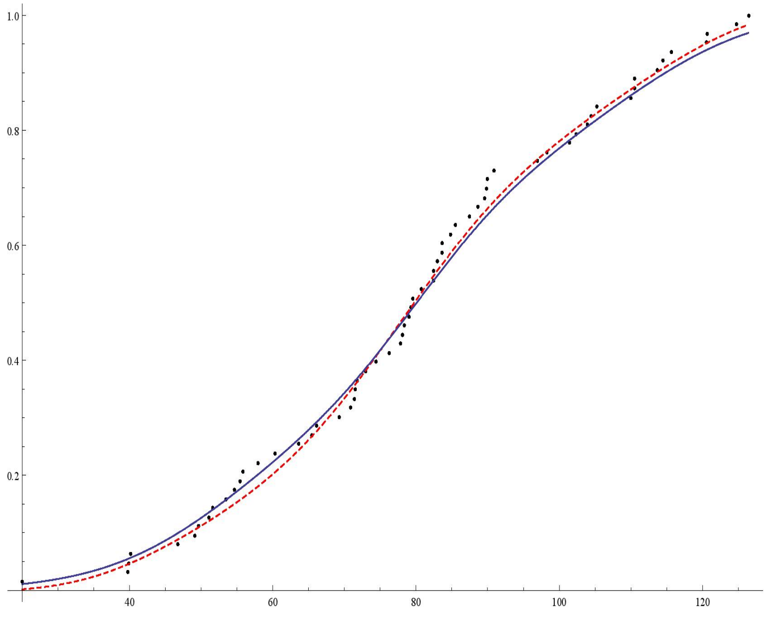

Figure 4.2: Saddlepoint CDF estimate adjusted with a 10^{th} degree polynomial (dashed line) and kernel CDF estimate (solid line) superimposed on the empirical CDF of the Buffalo snowfall data.

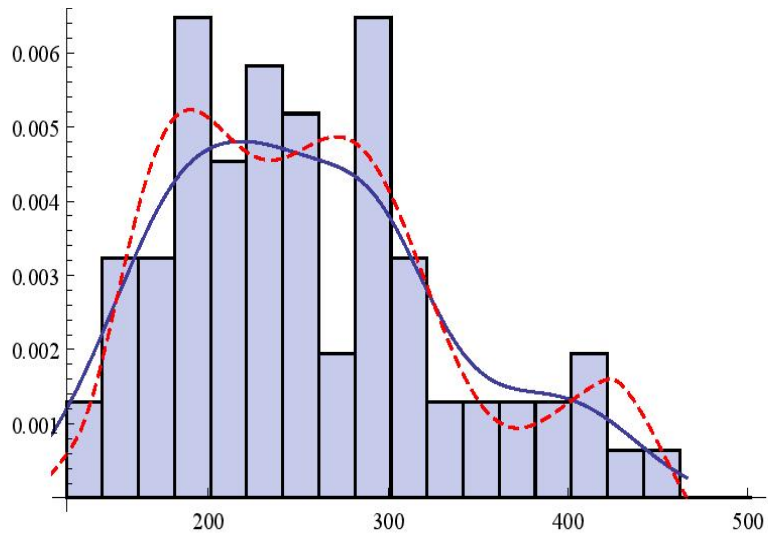

Figure 4.3: Adjusted saddlepoint density estimate (dashed line) and kernel density estimate (solid line) superimposed on a histogram of the flood peaks data.

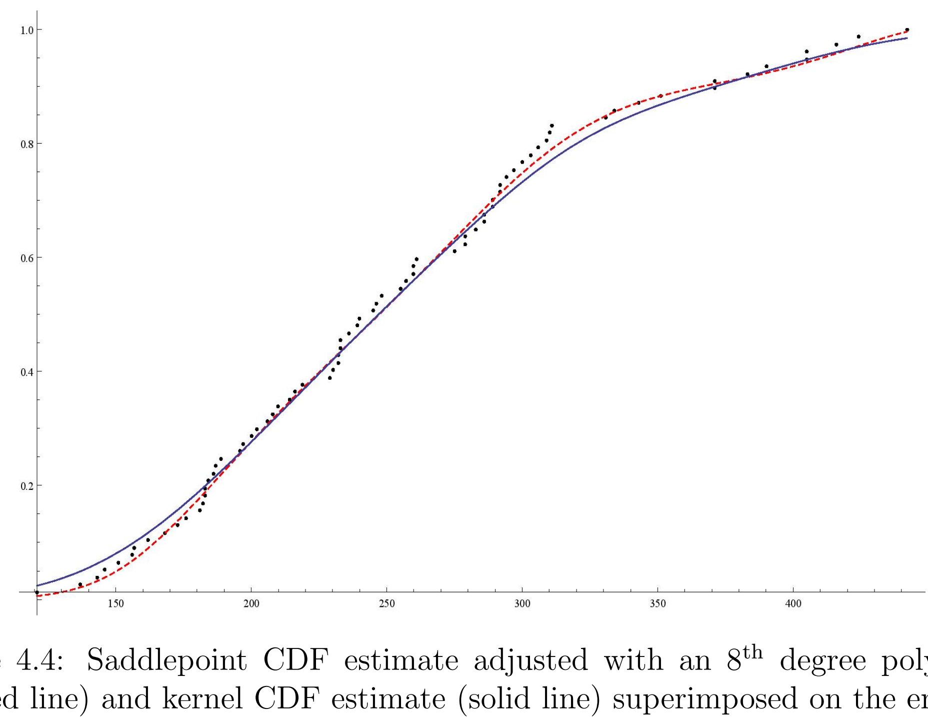

Figure 4.4: Saddlepoint CDF estimate adjusted with an 8^{th} degree polynomial (dashed line) and kernel CDF estimate (solid line) superimposed on the empirical CDF of the flood peaks data.

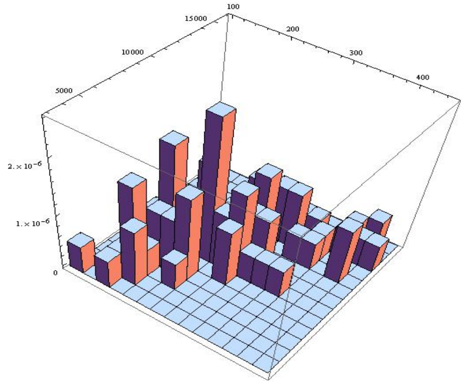

Figure 4.5: Histogram of the bivariate flood data.

4.2 Bivariate Applications

Flood Peaks and Volumes

Consider the bivariate flood data analyzed by Yue (2001), which includes observations on both peaks and volumes. A histogram of this data set appears in Figure 4.5. The bivariate density estimation methodology that is described in Section 3 was applied in conjunction with the sample estimates of the mean vector, covariance matrix and moments as well as the empirical cumulant-generating function associated with each of the standardized variables. Saddlepoint density estimates were determined for each variable in terms of normalized splines. Then, the bivariate polynomial adjustment defined in Equation (3.5), whose coefficients are obtained by solving Equation (3.7) was applied to the product of the marginal density estimates. A bivariate density estimate for the joint distribution of the original variables was finally obtained by applying the inverse transformation as specified by Equation (3.3).

The steps of the procedure are illustrated graphically in Figures 4.6–4.8. The product of the marginal densities of the standardized variables is plotted in Figure 4.6 and the joint density estimate of the standardized variables once adjusted with a bivariate polynomial of order 6 in each variable is plotted in Figure 4.7. It is seen from Figure 4.8 that, on applying the inverse of the standardizing transformation, the resulting adjusted saddlepoint density estimate appears to be more representative of the distributional features of the original data than the kernel density estimate shown in Figure 4.9.

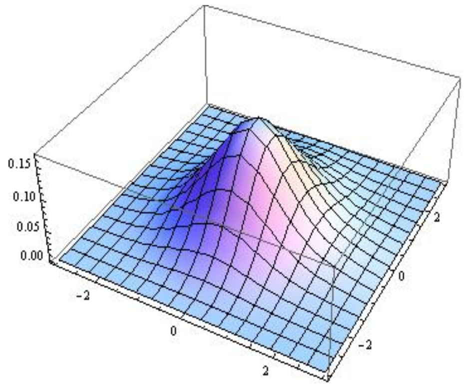

Figure 4.6: Product of the marginal density estimates of the standardized flood data.

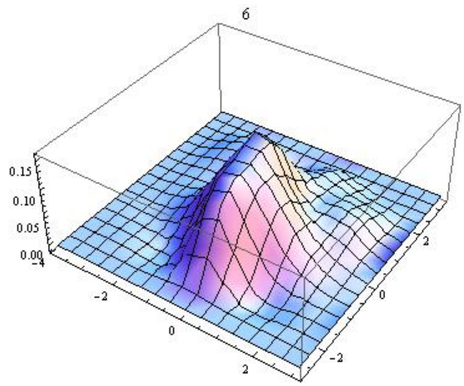

Figure 4.7: Adjusted density estimate of the standardized flood data ($t = 6$).

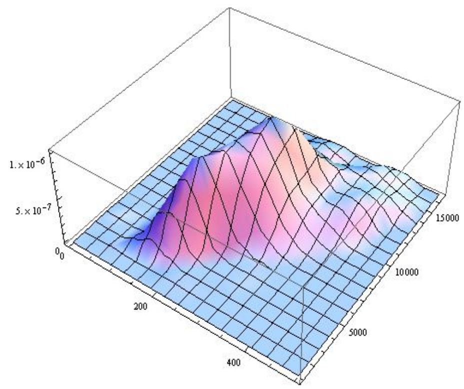

Figure 4.8: Adjusted saddlepoint density estimate of the original flood data after applying the inverse transformation.

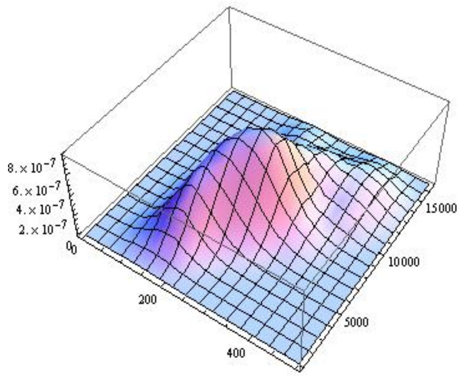

Figure 4.9: Kernel density estimate of the flood data.

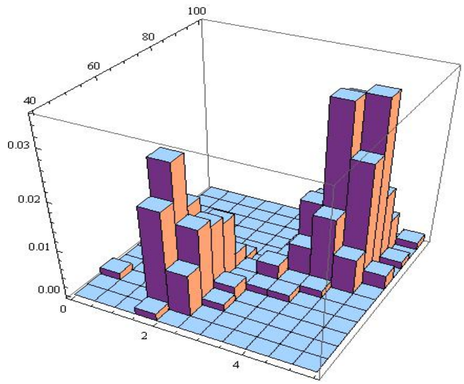

Figure 4.10: Histogram of the Old Faithful data.

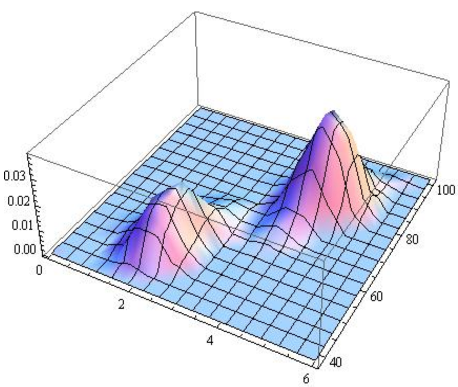

Figure 4.11: Adjusted saddlepoint density estimate of the Old Faithful data.

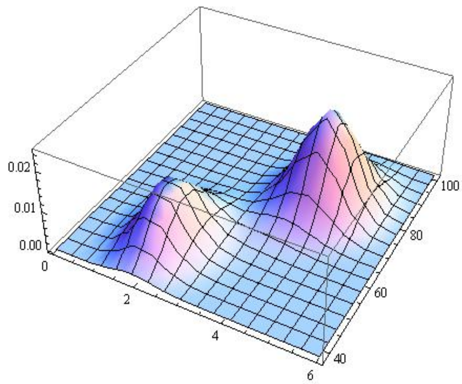

Figure 4.12: Kernel density estimate of the Old Faithful data.

The Old Faithful Data Set

The Old Faithful geyser data is available from the *R* package *datasets* (see also Azzalini and Bowman (1990)). This data set consists of the observed waiting times between eruptions and the duration of the eruptions in minutes. A histogram of the 272 bivariate observations is shown in Figure 4.10. On applying the proposed bivariate density estimation methodology, one obtains the adjusted saddlepoint density estimate shown in Figure 4.11. The kernel density estimate appearing in Figure 4.12 turns out to be to be very similar. All the calculations were carried out with the symbolic computation package *Mathematica*, the code being available from the authors upon request.

Acknowledgements

The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

References

[1] Azzalini, A. and Bowman, A. W. (1990). A look at some data on the Old Faithful geyser. *Applied Statistics* **39**, 357–365.

[2] Butler, R. W. (2010). *Saddlepoint Approximations with Applications*, Cambridge Series in Statistical and Probabilistic Mathematics, Book 22, Cambridge University Press.

[3] Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics* **25**, 631–650.

[4] Demaso, S. J., Hernandez, F., Brennan, L. A. and Bingham, R. L. (2011). Application of the simple saddlepoint approximation to estimate probability distributions in wildlife research. *The Journal of Wildlife Management* **75**(3), 740–746.

[5] Feuerverger, A. (1989). On the empirical saddlepoint approximation. *Biometrika* **76**(3), 457–464.

[6] Hu, Z., Ma, C. and Robinson, J. (2008). Empirical saddelpoint approximation of the standertized mean under stratified random sampling. *Statistics and Probability Letters* **28**, 396–401.

[7] Huzurbazar, S. (1999). Practical saddlepoint approximations. *The American Statistician* **53**, 225–232.

[8] Hyrien, O., Chen, R., Mayer-Proschel, M., and Noble, M. (2010). Saddlepoint approximations to the moments of multitype age-dependent branching processes, with applications. *Biometrics* **66**(2), 567–577.

[9] Jiang, M. and Provost, S. B. (2011). Improved orthogonal polynomial density estimates. *The Journal of Statistical Computation and Simulation* **81**(11), 1495–1516.

[10] Kolassa, J. and Li, J. (2010). Multivariate saddlepoint approximations in tail probability and conditional inference. *Bernoulli* **16**(4), 1191–1207.

[11] Kolassa, J. and Robinson, J. (2011). Saddlepoint approximations for likelihood ratio like statistics with applications to permutation tests. *The Annals of Statistics* **39**(6), 3357–3368.

[12] Lugannani, R. and Rice, S. O. (1980). Saddlepoint approximations for the distribution of sums of independent random variables. *Advances in Applied Probability* **12**, 475–490.

[13] Marsh, P. (2011). Saddlepoint and estimated saddlepoint approximations for optimal unit root tests. *Bootstrap and Numerical Methods in Time Series* **27**(5), 1026–1047.

[14] Provost, S. B. (2005). Moment-based density approximants. *The Mathematica Journal* **9**, 727–756.

[15] Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statistical Science* **3**, 213–238.

[16] Ronchetti, E. and Welsh, A. H.(1988). Empirical saddlepoint approximations for multivariate M-estimators. *Journal of the Royal Statistical Society. B, Statistical Methodology* **56**(2), 313–326.

[17] Smyth, G. K. and Podlich, H. M. (2002). An improved saddlepoint approximation based on the negative binomial distribution for the general birth process. *Computational Statistics* **17**, 17–28.

[18] Yue, S. (2001). A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrological Processes* **15**, 1033–1045.