*Bull. Cal. Math. Soc.*, **96**, (2) 139–154 (2004)

# ON THE CONSTRUCTIONS OF PROBABILITY DISTRIBUTIONS FOR DIRECTIONAL DATA[1]

ASHIS SENGUPTA

(*Received 20 September 2003*)

**Abstract.** Even though the support is finite, because of the disparity of the topologies between the circle and the line and as also of the additional requirement of periodicity of the density functions, as interesting situation is posed for the constructions of probability distributions for directional data. Drawing from certain results on characterization theory based on the calculus of variations and on functional annalysis, we enhance here the desirable properties of maximum entropy and conditional specifications. It is demonstrated in this paper how the adoption of such principles leads to the construction of interesting and useful probability distributions of the circle, torus and cylinder and their generalizations.

**1. Introduction.** Directional data arise in many diverse scientific investigations encountered frequently in our day-to-day life as observations on directions, orentations, angular displacements, periodic occurrences, etc. Construction of such distributions pose interesting problems since methods for linear data cannot be directly adopted due to the disparity of the tpoologies on the line and the circle. Further, the additional requirement of periodicity of such distributions need to be met. Several distributions have been proposed (Jammalamadaka and Sen Gupta, 2001, Chapter 2) to model such data, mostly on the circle. Here an attempt is made to unify certain results on characterizations of distributions and generalize them to yield not only new probability distributions on the circle but also those for the bivariate cases, such as for models on the torus, cylinder and their multivariate generaliztions.

**2. Maximum entropy characterizations.** The concept of information, equivalently of negative entropy, is commonly used in statistics. We demonstrate in this section how probability distributions may be characterized, and hence constructed, for directional data by invoking the criterion of maximization of certain entropy measures. While Shannon's entropy (section 2.1) has been the most popular one for such purposes, it is observed that maximization of certain other such functionals (section 2.2) also lead to useful probability distributions on the circle. A generalization of the basic result for the univariate case is seen to yield distributions on the torus, cylinder (section 2.3) and their multivariate extensions (section 4.1).

**2.1 Univariate case : Characterizations based on Shannon's entropy.** Let $X$ be a continuous random variable having the probability density function $f(x;\eta)$. Then Shannon's

[1]This paper is based in part on the Invited lecture presented by the author at the International Conference of the Calcutta Mathematical Society, Kolkata, December 2002.

information is defined as the negative entropy

$$H(f) = -\int_{-\infty}^{\infty} f(x;\eta)\log f(x;\eta)dx.$$

It follows, e.g., by results on majorization with Schur-convex functions (Marshall and Olkin, 1979), that this information measure is non-negative. Let

$$f(x;\eta) > 0, \ \ x \in (a,b) \ \text{and}, \ \ = 0 \ \text{otherwise}. \tag{1}$$

Consider the class $F$ of parametric density functions $f(x,\eta)$ that satisfy the constraints

$$\int_a^b T_j(x) f(x;\eta)dx = \tau_j, \ \ j = 1, 2, \cdots, k, \tag{2}$$

for a given set of integrable functions $T_1(x), \cdots, T_k(x)$ on $(a,b)$ and constraints $\tau_1, \cdots, \tau_k$. Then variational methods can be used to find the class of densities that maximize $H(f)$ over the class $F$. Using the Lagrangian,

$$L = -y\log y + y\sum_{i=1}^{k} \eta_i T_i(x)$$

and the corresponding Euler-Lagrange equation, the extremal density functions is given by :

THEOREM 1. *The maximum entropy over the class $F$ is attained by the exponential family of distributions, i.e., with the density of the form*

$$f(x;\eta) = C.\exp\left[\sum \eta_i \tau_i(x)\right], \tag{3}$$

*where $C$ is the normalizing constant to be determined by invoking the constraints* (1) *and* (2), *if (and only then) there exist $\eta_1, \eta_2, \cdots, \eta_k$ such that* (3) *satisfies the conditions* (1) *and* (2).

An alternative proof of the above characterization result based on inequalities is available from Theorem 13.2.1 of Kagan, Linnik and Rao (1973).

For distributions on the circle, it is customary to represent the random variable $X$ above by $\theta$ and the support by usually $[0, 2\pi)$ or by $[-\pi, \pi)$.

**2.1.1 Examples. 1. Circular Normal (CN) Distribution.** A circular r.v. $\theta$ is said to have a von Mises or a Circular Normal (CN) distribution (von Mises, 1918) if it has the probability density function :

$$f(\theta;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa\cos(\theta-\mu)}, \ \ 0 \le \theta < 2\pi, \tag{4}$$

where $0 \le \mu < 2\pi$ and $\kappa \ge 0$ are parameters. Here $I_0(\kappa)$, the normalizing constant, is the modified Bessel function of the first kind and order zero.

By taking $\tau_1$ and $\tau_2$ consistent with the expectations of $T_1(\theta) = \cos(\theta)$ and $T_2(\theta) = \sin(\theta)$ respectively, i.e., by specifying the first harmonic of $\theta$, (3) yields (4).

**2. Multimodal CN distributions.** A circular r.v. $\theta$ is said to have a $p$-modal von Mises or a $p$-modal CN distribution ($p$ known), if it has the probability density function :

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos p(\theta-\mu)}, \quad 0 \leq \theta < 2\pi, \tag{5}$$

where $0 \leq \mu < 2\pi$ and $\kappa \geq 0$ are parameters.

By taking $\tau_1$ and $\tau_2$ consistent with the expectations of $T_1(\theta) = \cos p(\theta)$ and $T_2(\theta) = \sin p(\theta)$ respectively, i.e., by specifying the $p$-th harmonics of $\theta$, (3) yields (5).

**3. A skewed CN distribution.** A three parameter circular distribution which is a member of the exponential family and may be looked upon as a generalization of the two parameter CN distribution, has been suggested by Rattihalli and Sen Gupta (2002). This density (to be referred to as the SGR density) is given by

$$f(\theta; \mu_1, \kappa_1, \kappa_2) = C.\exp(\kappa_1 \cos(\theta - \mu_1) + \kappa_2 \cos 2\theta),$$
$$0 \leq \theta < 2\pi, \quad 0 \leq \mu_1 < 2\pi, \quad \kappa_1, \kappa_2 \geq 0, \tag{6}$$

where $C$ is the normalizing constant and can be expressed in terms of the weighted sum of independent non-central $\chi^2$ variables or may be computed in the same lines as done for example 4 below.

By taking $\tau_1, \tau_2$ and $\tau_3$ consistent with the expectations of $T_1(\theta) = \cos(\theta)$, $T_2(\theta) = \sin(\theta)$ and $T_3(\theta) = \cos 2(\theta)$ respectively, (3) yields (6).

**4. Rukhin's generalized CN distribution.** A four parameter circular distribution which is a member of the exponential family and may be looked upon as yet another generalization of the two parameter CN distribution, has been suggested by several researchers, e.g., Rukhin (1972), Cox (Mardia, 1975a), Beran (1979). This density is given by,

$$f(\theta; \alpha, \beta, a, b) = C.\exp(a \cos(\theta - \alpha) + b \cos 2(\theta - \beta)),$$
$$0 \leq \theta < 2\pi, \quad 0 \leq \alpha, \quad \beta < 2\pi, a \geq 0, b \geq 0, \tag{7}$$

where $C$ is the normalizing constant, which may be obtained as below Yfantiz and Borgman, 1982). After calculating the Fourier series expansions of

$$F_1 = e^{a \cos(\theta-\alpha)}, \quad \text{and} \quad F_2 = e^{b \cos(2(\theta-\beta))},$$

and using some manipulations, it turns out that,

$$C^{-1} = 2\pi \left\{ I_0(a) I_0(b) + 2 \sum_{n=1}^{\infty} I_n(b) I_{2n}(a) \cos(2n(\beta - \alpha)) \right\},$$

here $I_n(a)$, $I_n(b)$ are the modified Bessel functions of the first kind and order $n$, i.e.,

$$I_n(s) = \frac{1}{\pi}\int_0^{\pi} e^{s\cos\theta}\cos(n\theta)d\theta, \quad s = a, b.$$

By taking $\tau_1, \tau_2, \tau_3$ and $\tau_4$ consistent with the exprectations of $T_1(\theta) = \cos(\theta)$, $T_2(\theta) = \sin(\theta)$, $T_3(\theta) = \cos 2(\theta)$ and $T_4(\theta) = \sin 2(\theta)$ respectively for the above distribution, (3) yields (7).

Both the distributions in (6) and (7) are capable of capturing a variety of shapres, both symmetric and asymmetric, both unimodal and bimodal. However, note that (6) requires one parameter less than (7).

**2.2 Univariate distributions derived from other entropy measures/functionals.** We now briefly discuss the invocation of other entropy measures for deriving circular distributions. Solutions (Ochoa and Delgado-Gonzalez, 1990) through variational trechniques may be obtained to the classic inverse theory problem arising out of characterizing a density function $D(\theta)$ which under isoperimetric constraints minimizes (maximizes) a relevant combination of some integral measure $\psi$ of $D$ or a functional (an entropy in most cases) for various such functional. $D(\theta)$ is obtained by minimizing a fairly general functional,

$$\psi = \int_{-\pi}^{\pi}\left[F(D) + \{\lambda_0 + \lambda_1\cos\theta + \lambda_2\sin\theta + \lambda_3\cos 2\theta + \lambda_4\sin 2\theta\}D(\theta)\right]d\theta$$

Note that our constrains correspond to specifying the first two harmonics of $D$ (first two trigonometric moments of $\theta$). Unimodal versions of the solutions to various choices of $F(D)$ yield various circular probability distributions. Below we show how some particular such choices yield familiar circular distributions.

**2.2.1 Examples. General Wrapped Stable (WS) family of distributions.** Wrapped distributions on the circle are obtained by wrapping the corresponding distribution on the line, i.e., by using $\theta = X \bmod 2\pi$. Note that the trinonometric moments uniquely characterize a circular distribution. Further, the trigonometric moment of order $p$ for a wrapped circular variable $\theta$ corresponds to the value of the characteristic function of the unwrapped linear random variable $X$, say $\phi_x(t)$ at the integer value $p$, i.e., $\phi_p = \phi_x(p)$ (Proposition 2.1 of Jammalamadaka and Sen Gupta, 2001). Thus wrapped $\alpha$-stable distributions may be constructed by using the characteristic function of the $\alpha$-stable distribution of the real line, which is given by

$$\varphi(t) = \begin{Bmatrix} \exp\left\{-\tau^{\alpha}|t|^{\alpha}\left[1 - i\beta\,\mathrm{sgn}\,(t)\tan\frac{\alpha\pi}{2}\right] + i\mu t\right\}, & \text{if } \alpha\varepsilon(0,1)\cup(1,2], \\ \exp\{-\tau|t| + i\mu t\}, & \text{if } \alpha = 1, \end{Bmatrix}$$

with $\tau \geq 0$, $|\beta| \leq 1, 0 < \alpha \leq 2$, while $\mu$ is a real number.

The density function of a wrapped $\alpha$-stable random variable $\theta \in [0, 2\pi)$, is given by (Lukacs, 1970)

$$f(\theta) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty}\exp\{-\tau^{\alpha}k^{\alpha}\}\cos\left\{k(\theta-\mu) - \tau^{\alpha}k^{\alpha}\beta\tan\frac{\alpha\pi}{2}\right\},$$

where $\alpha \in (0,1) \cup (1,2]$, with $\mu$ conveniently redefined as $\mu = \mu \pmod{2\pi}$. Note that although there is generally no closed form expression for the density of an $\alpha$-stable distribution on the real line, we are able to write such density for the wrapped case, at least as an infinite series. The particular case corresponding to $\beta = 0$ gives us the symmetric Wrapped Stable (SWS) family of circular densities, to be simply referred to as Wrapped Stable (WS) family, given by

$$f(\theta; \rho, \alpha, \mu) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \rho^{k^{\alpha}} \cos\{k(\theta - \mu)\}, \tag{8}$$

where $\rho = \exp(-\tau^{\alpha})$. For further discussions see, e.g., Jammalamadaka and Sen Gupta (2001).

When $\rho = 0$, we get the Circular Uniform (CU) distribution, $\alpha = 1$, and 2 give the Wrapped Cauchy (WC) (Levy, 1939) and wrapped normal (WN) distributions, respectively. Further, if only the first term in the summation is retained, one gets the Cardioid distribution.

Observe now that optimal solutions to maximizing the functional $\Psi$ for various choices of $F(D)$ yield familiar circular distributions. For example, $F(D) = -\ln D$ yields the WC density (Lygre and Krogstad, 1986); $F(D) = -D \ln D$ yields the CN or von Mises density (Burg, 1975) – recall (Collett and Lewis, 1981) that any CN distribution may be represented for all practical purposes by the corresponding WN distribution; $F(D) = D^2$ yields the Cardioid density (Long and Hasselmann, 1979). Other appropriate choices of $F(D)$ will yield new circular probability distributions. However, the identification of the general functional $F(D)$ with corresponding suitable constraints which will yield, at least, the symmetric wrapped stable family is still an open problem.

Use of distributions from the WS family enables one to capture higher kurtosis in the data which is otherwise not possible through the CN distribution. These distributions have been found useful (Beal, 1991; Tucker, 1991) in the context of several aspects of wave theory. From both theoretical and parametric "complete class" of useful unimodal distributions for circular data.

The shapes of the various distributions discussed above, as displayed over the circle, may be viewed through the statistical package DDSTAP developed by the author (Sen Gupta, 1998).

**2.3 Bivariate Case : Characterizations based on Generalization of Theorem 1.** Suppose that we are interested in distributions defined over a space $S$ and that these distributions are to be represented by densities relative to some familiar measure such as Lebsegue, Harr, etc. let $t_1, \cdots, t_q$ represent specified $q$ real valued measurable functions over $S$ such that no linear combination of $t_i, \cdots, t_q$ is constant.

THEOREM 2. *If for a probability density function* $f(x)$

(i) $S_1$ *is the support of* $f(x)$ *where* $x \in S_1, S_1 \subset S$,

(ii) $E\{t_i(x)\} = a_i$ *(fixed)*, $i = 1, \cdots, q$,

(iii) *the entropy is maximized,*

*then* $f(x)$ *should be of the form*

$$f(x) = \exp\left\{b_0 + \sum_{i=1}^{q} b_i t_i(x)\right\}, \quad x \in S_1, \tag{9}$$

*provided there exist* $b_0, b_1, \cdots,$ *such that* (9) *satisfies* (i) *and* (ii). *Further, if there exists such a density, then it is unique.*

*Proof:* See Theorem 3.1 of Mardia (1975b).

**2.3.1 Examples. 1. Distributions on the Torus.** Let $\underset{\sim i}{\ell} \in S_2$, $i = 1, 2$. Let $E(\underset{\sim 1}{\ell}), E(\underset{\sim 2}{\ell})$ and $E(\underset{\sim 1}{\ell'}\underset{\sim 2}{\ell})$ be specified. From (9) this yields the distribution on the torus given by (Mardia, 1975b).

$$f(\theta,\phi) = C.\exp[\kappa_1 \cos(\theta-\mu) + \kappa_2 \cos(\phi-\nu) + \rho(\kappa_1\kappa_2)^{1/2}\cos(\theta+\phi-\psi)], \tag{10}$$

where, $\ell_1 = (\cos\theta, \sin\theta), \ell_2 = (\cos\phi, \sin\phi)$, $0 \le \theta,\ \psi < 2\pi, \kappa_1, \kappa_2 > 0$, $0 \le \mu, \nu, \psi < 2\pi$ and

$$C = I_0(\kappa_1)I_0(\kappa_2)I_0\left\{\rho(\kappa_1\kappa_2)^{1/2}\right\} + 2\sum_{p=1}^{\infty} I_p(\kappa_1)I_p(\kappa_2)I_p\left\{\rho(\kappa_1\kappa_2)^{1/2}\right\}\cos p\psi.$$

Using the same constraints for the marginals but invoking the constraint for a joint moment as $E(\sin(\theta-\mu)\sin(\phi-\nu))$ being specified, Jammalamadaka and SenGupta (2001) obtained another distribution on the torus given by

$$f(\theta,\phi) = C.\exp\left[\kappa_1\cos(\theta-\mu) + \kappa_2\cos(\phi-\nu) + \kappa_3\sin(\theta-\mu)\sin(\phi-\nu)\right], \tag{11}$$

where $C$ is the normalizing constant.

This distribution imbeds the bivariate normal distribution with a small range for the observations, permitting the quadratic and linear terms of the latter to be replaced by their circular analogues. The distribution in (11) has been used by Singh et al. (2002) for probabilistic modelling of torsional angles in molecules.

Recently, Arnold and SenGupta (2002) has suggested the distribution given in (15). Because of its genesis as discussed in section 3.2, its functional form and other details are given in that section. However, we note that this distribution corresponds to the maximum entropy distribution obtained on specifying the first marginal and first joint harmonics of $\theta$ and $\phi$, i.e., on specifying $E(\cos\theta)$, $E(\sin\theta)$, $E(\cos\phi)$, $E(\sin\phi)$ and $E(\cos\theta\cdot\cos\phi)$, $E(\cos\theta\cdot\sin\phi)$, $E(\sin\theta\cdot\cos\phi)$, $E(\sin\theta\cdot\sin\phi)$.

**2. Distributions on the Cylinder.** Since the supports for the random variables in theorem 2 are quite general, the joint distribution of a linear random variable and a circular random variable may be derived from it using suitable constraints. This enabled Johnson and Wehrly (1978) to present a variety of distributions, as given below, on the cylinder.

**(2.1).** The density function of $(\theta, X)$ given by

$$f(\theta, x) = (\lambda^2 - \kappa^2)^{1/2}(2\pi)^{-1}.\exp\{-\lambda x + \kappa x \cos(\theta - \mu)\},$$

where $0 \leq \theta, 2\pi$, $x > 0$, $0 < \kappa < \lambda$, and $0 \leq \mu < 2\pi$, is the maximum entropy distribution subject to $E(X)$, $E(X\cos\theta)$, and $E(X\sin\theta)$ taking specified values which are consistent with expectation with respect to the above distribution.

**(2.2)** Let $(\theta, X)$ have the joint density

$$f(\theta, x) = C\cdot\exp\left\{-\frac{x^2}{2\sigma^2} + \frac{\lambda x}{\sigma^2} + \frac{\kappa x}{\rho^2}\cos(\theta - \mu)\right\},$$

where $C > 0$ is a constant of integration, $-\infty < x < \infty$, $0 \leq \theta < 2\pi$, $-\infty < \lambda < \infty$, $\kappa, \sigma > 0$ and $0 \leq \mu < 2\pi$. Then $f(\theta, x)$ is the maximum entropy angular-linear distribution subject to $E(X), E(X^2), E(X\cos\theta)$ and $E(X\sin\theta)$ taking specified values consistent with the expectations with respect to the above distribution.

**(2.3).** Let $(\theta, X)$ have the joint density

$$f(\theta, x) = C\cdot\exp\{-\lambda x + \kappa x\cos(\theta - \mu_1) + \nu\cos(\theta - \mu_2)\},$$

where $0 \leq \theta < 2\pi$, $0 < x < \infty$, $\lambda > \kappa > 0$, $0 \leq \mu_1, \mu_2 < 2\pi$, and

$$C = (\lambda^2 - \kappa^2)^{1/2}(2\pi)^{-1}\cdot\left\{I_0(\nu) = 2\sum_{p=1}^{\infty}\rho^p I_p(\nu)\cos[p(\mu_1 - \mu_2)]\right\}^{-1},$$

$$\rho = \kappa[\lambda + (\lambda^2 - \kappa^2)^{1/2}]^{-1},$$

with $I_p(\cdot)$ being the modified Bessel function of the first kind and other $p$. Then $f(\theta, x)$ is a maximum entropy density subject to $E(X)$, $E(\cos\theta)$, $E(\sin\theta)$, $E(X\cos\theta)$ and $E(X\sin\theta)$ taking specified values consistent with expectation with respect to the above.

The proofs of the above reuslts 1–3 are quite straightforward on noting theorem 2.

**3. Specified Conditionals Family Characterizations.** Often it is desirable to specify the conditionals of a multivariate distribution. Arnold and Strauss (1991) gave an unified approach of characterizing the class of bivariate distributions such that the conditional distributions belong to any specified exponential family. Even though they considered linear vector variables, we show below that their results can be exploited to yield also distributions on the torus, cylinder (section 3.1) and their generalizations (section 4.2).

**3.1 Bivariate case : Characterizations based on conditionals exponential family specifications**

THEOREM 3. *All solutions of the equation*

$$\sum_{k=1}^{n} f_k(x)g_k(y) = 0, \quad x \in S(X), \quad y \in S(Y),$$

*can be written in the form*

$$\begin{pmatrix} f_1(x) \\ f_2(x) \\ \cdots \\ f_n(x) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nr} \end{pmatrix} \begin{pmatrix} \Phi_1(x) \\ \Phi_2(x) \\ \cdots \\ \Phi_r(x) \end{pmatrix}$$

$$\begin{pmatrix} g_1(y) \\ g_2(y) \\ \cdots \\ g_n(y) \end{pmatrix} = \begin{pmatrix} b_{1r+1} & b_{1r+2} & \cdots & b_{1n} \\ b_{2n+1} & b_{2r+2} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \\ \cdots & & & \\ b_{nr+1} & b_{nr+2} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} \Psi_{r+1}(y) \\ \Psi_{r+2}(y) \\ \cdots \\ \Psi_n(y) \end{pmatrix}$$

*where $r$ is an integer between 0 and $n$, and $\Phi_1(x)$, $\Phi_2(x), \cdots, \Phi_r(x)$ on the one hand and $\Psi_{r+1}(x), \Psi_{r+2}(x), \cdots, \Psi_n(x)$ on the other are arbitrary systems of mutually linearly independent functions and the constants $a_{ij}$ and $b_{ij}$ satisfy*

$$\begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1r} & a_{2r} & \cdots & a_{nr} \end{pmatrix} \begin{pmatrix} b_{1r+1} & b_{1r+2} & \cdots & b_{1n} \\ b_{2r+1} & b_{2r+2} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{nr+1} & b_{nr+2} & \cdots & b_{nn} \end{pmatrix} = 0$$

Theorem 3 in the well known Aczel's theorem. As a simple corollary to it, we have the following (Arnold et al. 1999).

COROLLARY. *All solutions of the equation*

$$\sum_{r=1}^{r} f_i(x)\Phi_i(y) = \sum_{j=1}^{s} g_j(y)\Psi_j(x), \quad z \in S(X), \ y \in S(Y),$$

*where $\{\Phi_i\}_{i=1}^{r}$ and $\{\Psi_j\}_{j=1}^{s}$ are given systems of mutually linearly independent functions, are of the form*

$$\underline{f}(x) = C\underline{\Psi}(x) \ \text{ and } \ \underline{g}(y) = D\underline{\Psi}(y), \ \text{ where } D = C'.$$

For constructing joint distributions with conditionals being specified, the above results are exploited to yield the fundamental theorem.

THEOREM 4. (Arnold and Strauss, 1991) *Let $f_1(x;\eta)$ and $f_2(y;\tau)$ denote members of two $\ell_1$- and $\ell_2$-parameter exponential families. Let $f(x,y)$ be a bivariate density whose conditional densities satisfy*

$$f(x|y) = f_1(x;\underline{\eta}(y)) \tag{12}$$

*and*

$$f(y|x) = f_2(y;\underline{\tau}(x)) \tag{13}$$

*for some function $\underline{\eta}(y)$ and $\underline{\tau}(x)$. It follows that $f(x,y)$ is of the form*

$$f(x,y) = r_1(x)r_2(y)\exp\{\underline{q}^{(1)}(x)'M\underline{q}^{(2)}(y)\} \tag{14}$$

*where*

$$\begin{aligned}\underline{q}^{(1)}(x) &= (q_{10}(x), q_{11}(x), q_{12}(x), \cdots, q_1\ell_1(x)),\\ \underline{q}^{(2)}(y) &= (q_{20}(y), q_{21}(y), q_{22}(y), \cdots, q_2\ell_2(y)).\end{aligned}$$

where $q_{10}(x) = q_{20}(y) \equiv 1$ and $M$ is a matrix of constants parameters of approximate dimensions (*i.e.*, $(\ell_1+1)\times(\ell_2+1)$) subject to the requirement that

$$\int_{D_1}\int_{D_2} f(x,y)d\mu_1(x)d\mu_2(y) = 1.$$

For convenience we can partition the matrix $M$ as follows :

$$M = \left(\begin{array}{c|cccc} m_{00} & m_{01} & \cdots & m_{0\ell_2} \\ \hline m_{10} & & & \\ \cdots & & \tilde{M} & \\ m_{\ell_1 0} & & & \end{array}\right)$$

Note that the case of independence is included through the choice $\tilde{M} = 0$.

**3.2 Examples.** Even though Arnold and Strauss (1991) used theorem 4 to construct bivariate distributions for only linear random variables, it is easy to see that the result extends to a general vector variable with both or one of its components being circular. Thus, through this extension, bivariate distributions on both the torus and cylinder may be derived.

**1. Distributions on the Torus.** Consider first bivariate distributions on the torus with CN conditionals (CNC). Since CN distributions are members of the exponential family, theorem 4 readily applies to yield the CNC distribution (Arnold and SenGupta, 2002)

$$f(\theta,\phi) = \exp[p'(\theta)Mq(\phi)], \quad (\theta,\phi)\in[0,2\pi)^2 \tag{15}$$

where

$$p(\theta) = \begin{pmatrix} 1 \\ \cos\theta \\ \sin\theta \end{pmatrix},$$

$$q(\phi) = \begin{pmatrix} 1 \\ \cos\phi \\ \sin\phi \end{pmatrix},$$

$$\text{and } M = \begin{pmatrix} m_{00} & m_{01} & m_{02} \\ m_{10} & m_{11} & m_{12} \\ m_{20} & m_{21} & m_{22} \end{pmatrix}$$

$M$ is the matrix of parameters with $m_{00}$, a function of the other $m_{ij}s$, being the normalizing constant.

The marginal densities are not CN densities unless independence holds—the same is true for the distributions on the torus obtained in section 2.3.1.

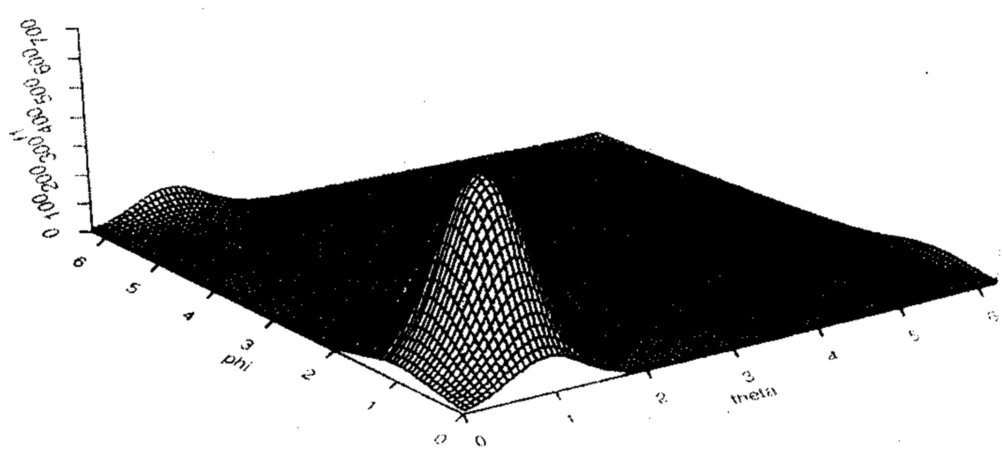


**Fig. 1.** Scaled pdf on the torus for $(m_{01}, m_{02}, m_{10}, m_{11}, m_{12}, m_{20}, m_{21}, m_{22} = (2, 3, 4, 0, 0, 5, 0, 0)$.

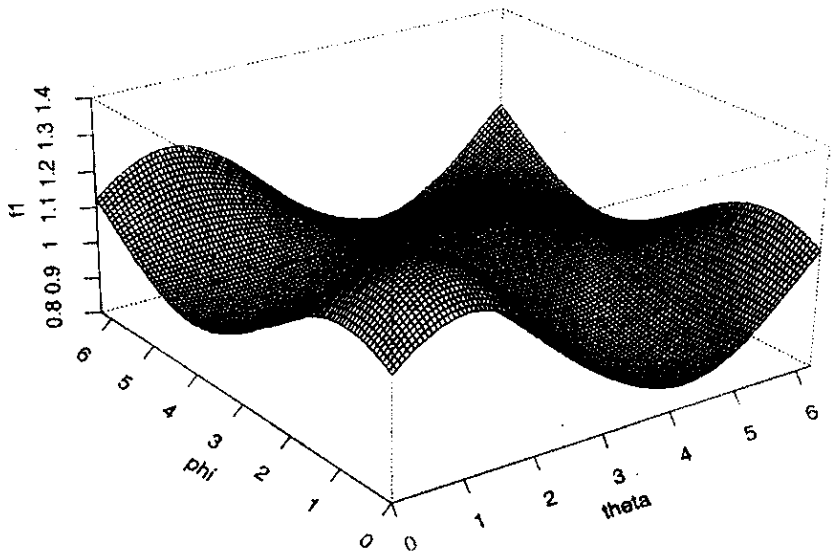


**Fig. 2.** Scaled pdf on the torus for $(m_{01}, m_{02}, m_{10}, m_{11}, m_{12}, m_{20}, m_{21}, m_{22} = (.02, .03, .04, .05, .06, .07..08, .09)$.

Figures 1 and 2 depict the shapes of two distributions on the torus with their corresponding parameter values, the former being for two independent circular variables. Note that both the distributions (10) and (11) are obtained as special cases of (15) by setting appropriately some of its parameters to be zeros.

**2. Distributions on the Cylinder.** Let,

$$f(\theta, x) = \exp\cdot [p'(\theta)Mq(x)], \quad (\theta, x) \in [0, 2\pi) \times R$$

where

$$p(\theta) = \begin{pmatrix} 1 \\ \cos\theta \\ \sin\theta \end{pmatrix} \quad \text{and} \quad q(x) = \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}.$$

Then $f(\theta, x)$ above defines the CN-normal (angular-linear) conditionals family of distributions. It is obvious from the above derivations that one may derive a spectrum of bivariate distributions on both the torus and the cylinder by this conditional approach. One may thus obtain CN-CN, CNRukhin, CN-SGR, SGR-Rukhin, etc. conditionals distributions on the torus. Similarly, CN-Exponential, CN-Beta, CN-Normal, Rukhin-Normal, SGR-Normal, etc. conditionals distributions may be easily derived for random variables jointly distributed on the cylinder. Exploiting the same theorem one may also construct joint distributions when the linear variable happens to be discrete, e.g., CN-Poisson, CN-Binomial, SGR-Poisson, Rukhin-Poisson, etc.

conditionals distributions. Figures 3 and 4 depict the shapes of two distributions on the cylinder with their corresponding parameter values, the former being for two independent linear and circular variables.

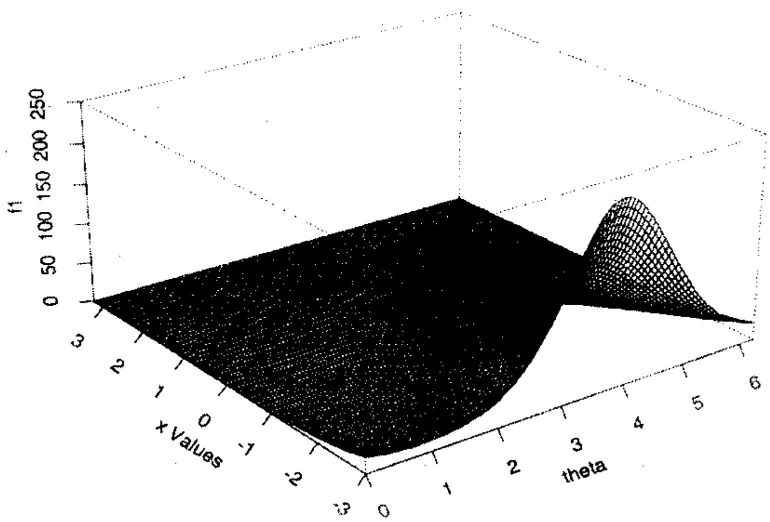


**Fig. 3.** Scaled pdf on the cylinder for $(m_{01}, m_{02}, m_{10}, m_{11}, m_{12}, m_{20}, m_{21}, m_{22} = (-1, 0, -1, 0, 0, 0, 0, 0)$.

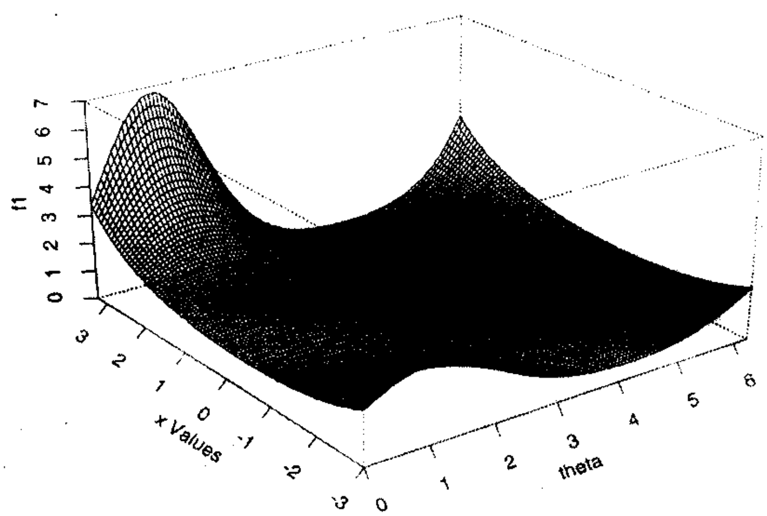


**Fig. 4.** Scaled pdf on the cylinder for $(m_{01}, m_{02}, m_{10}, m_{11}, m_{12}, m_{20}, m_{21}, m_{22} = (.02, .03, .04, .05, .06, .07, .08, .09)$.

**4. Multivariate case : Generalizations.** Both the approaches discussed above may be generalized to yield joint distributions for the multivariate case, i.e., when all or some of the variables are circular random variables. We discuss below these situations briefly. The generalized forms of the distributions may be written down with some care-however, the details are omitted since the notations become increasingly complex.

### 4.1 Maximum entropy distributions

#### 4.1.1 Examples

**1. Distributions on the Hypershere.** Let $\underset{\sim i}{\ell} \in S_p$, $i = 1, 2$, and let these two random vector variables be correlated. Its easiest to specify the first order marginal and joint moments, $E(\underset{\sim 1}{\ell})$, $E(\underset{\sim 2}{\ell})$ and $E(\underset{\sim 1}{\ell'}\underset{\sim 2}{\ell})$. In this case, the maximum entropy density from theorem 2 is

$$C \cdot \exp(\underset{\sim 1}{a'}\underset{\sim 1}{\ell} + \underset{\sim 2}{a'}\underset{\sim 2}{\ell} + \underset{\sim}{a'}\underset{\sim 1}{\ell}\underset{\sim 2}{\ell'}\underset{\sim}{b}), \tag{16}$$

where $C$ is the normalizing constant. The distribution given by (16) is termed (Mardia, 1975b) a generalized von Mises-Fisher Distribution.

For $p = 2$, we get the distribution on the torus as given in section 2.3.1.

**2. Distributions on the Hypercylinder.** Let,

$$H(\theta) = \begin{pmatrix} \cos\theta_1 & \cdots & \cos n\theta_1 & \sin\theta_1 & \cdots & \sin n\theta_1 \\ \cdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \cos\theta_p & \cdots & \cos n\theta_p & \sin\theta_p & \cdots & \sin n\theta_p \end{pmatrix}$$

Let $\Theta$ and $X$ have the joint density function

$$f(\theta, x) = C \cdot \exp\left\{-\frac{1}{2}x'\sum{}^{-1}x + \lambda'\sum{}^{-1}x + a(\theta)'\sum{}^{-1}x\right\}, \tag{17}$$

where $C$ is a constant of integration, $a(\theta)' = (a_1(\theta), \cdots, a - q(\theta))$,

$$\begin{aligned} a_i(\theta) &= \sum_{j=1}^{p}\sum_{k=1}^{n} a_{ijk}\cos[k(\theta_j - \mu_{ijk})] \\ &= \sum_{j=1}^{p}\sum_{k=1}^{n}[\alpha_{ijk}\cos(k\theta_j) + \beta_{ijk}\sin(k\theta_j)], \quad i = 1, \cdots, q, \end{aligned}$$

$x \in R^q$, $0 \in [0, 1\pi)^p$, and $\sum^{-1}$ is positive definite. Then (Johnson and Wehrly, 1978) $f(\theta, x)$ maximizes the entropy of multivariate angular-linear distributions subject to $E[XX']$, $E(X)$, and $E[X \otimes H(\Theta)]$, where $\otimes$ is the Kronecker product, taking specified values consistent with expectation with respect to the distribution (17).

**Note.** The conditional distribution of $X$ given $\Theta = \theta$ is $q$-dimensional multivaiate normal with mean $\lambda + a(\theta)$ and covariance matrix $\sum$. Consequently, this model leads to a natural method of predicting a linear vector variable $X$ from a circular vector variable $\theta$ of directions. The conditional distribution of $\Theta$ given $X = x$ may be related to a multivariate circular distribution with independent components. Being a member of the exponential family, this is easily seen for $n = 1$ to be a special case of the multivariate CN-multivariate normal conditionals distribution to be discussed in the next section.

**4.2 Multivariate conditionally specified distributions.** It is indicated in section 8 of Arnold and Strauss (1991) how the result in theorem 4 may be generalized to the case of $k > 2$ variables with the conditional density of each being a specified member of the exponential family. The resulting joint density is again a member of the exponential family. We extend this generalization to cover multivariate circular as well as linear-circular variables. Consider first the joint density of $k$ circular random variables $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)$, $\underline{\theta} \in [0, 2\pi)^k$. Let us characterize the class of densities for $\underline{\theta}$ such the conditional density of each $\theta_i$ given the rest is a CN density. Since the CN density is in the exponential family, it follows from Arnold and SenGupta (2002) that the family of joint densities is given by the $(3^k - 1)$ parameter exponential family of densities :

$$f_{\underline{\Theta}}(\underline{\theta}) = \exp\left\{\sum_{i_1=0}^{2}\sum_{i_2=0}^{2}\cdots\sum_{i_k=0}^{2} a_{i_1,i_2,\cdots,i_k}\left[\prod_{j=1}^{k} q_{i_j}(\theta_j)\right]\right\}, \quad \underline{\theta} \in [0, 2\pi)^k \tag{18}$$

where we have defined functions $q_0, q_1$ and $q_2$ as : $q_0(u) \equiv 1$, $q_1(u) = \cos u$, and $q_2(u) = \sin u$. For $k > 2$, setting some of the parameters equal to zero will yield submodels with a lower dimensional parameter space.

It is obvious from the above construction that, in lieu of the CN density, other circular densities in the exponential family, e.g., the $p$-modal CN, Rukhin or SGR densities, can be taken as the conditional densities to yield a spectrum of multivariate circular distributions which are all different members of the exponential family.

Now suppose that we are interested in deriving joint distributions of random variables defined over the hypertorus, i.e., when some of the variables are circular while the rest are linear. Let us impose the constraints that the conditional distribution of each variable (linear or circular) given the rest is specified as a member of the exponential family with parameters possibly being functions of the conditioning variables. It is then obvious from the derivation for the multivariate circular case above, that this joint density will be a member of the multivariate multiparameter exponential family and can be written down explicitly.

**Acknowledgement.** The author would like to thank Mr. Sabyasachi Bhattacharyay for his help with the figures in this paper.

## *References*

**Arnold, B.C., Castillo, E.** and **Sarabia, J.M.** (1999) : *Conditional Specification of Statistical Models.* Springer-Verlag, New York.

**Arnold, B.C.** and **Strauss, D.** (1991) : Bivariate distributions with conditionals in prescribed exponential families. *J. Roy. Statist. Soc. B*, **53**, 365.

**Arnold, B.C.** and **SenGupta, A.** (2002) : Distributions with circular normal conditionals and tests for independence on the torus. *Tech. Rep.* 273, University of California, Riverside.

**Beran, R.** (1979) : Exponential models for directional data. *Ann. Statist.*, **7**, 1162.

**Beal, R.C.** (Ed). (1991) : *Directional Ocean Wave Spectra.* The Johns Hopkins University Press, Maryland.

**Burg, J.P.** (1975) : *Maximum Entropy Spectral Analysis.* Ph.D. Thesis. Stanford University.

**Collett, D.** and **Lewis, T.** (1981) : Discriminating between the von Mises and wrapped normal distributions. *Aust. J. Statist.*, **23**, 73.

**Jammalamadaka, S.R.** and **SenGupta, A.** (2001) : *Topics in Circular Statistics.* World Scientific Publ., New Jersey.

**Johnson, R.A.** and **Wehrly, T.E.** (1978) : Some angular-linear distributions and related regression models. *J. Amer. Statist. Assocn.*, **73**, 602.

**Kagan, A.M., Linnik, Yu. V.** and **Rao, C.R.** (1973) : *Characterization problems in mathematical statistics*, J. Wiley, New York.

**Lévy, P.** (1939) : L'addition des variables aléatoires définies sur une circonférence. *Bull. Soc. Math. France*, **67**, 1.

**Long, R.B.** and **Hasselmann, K.** (1979) : A variational technique for extracting directional spectra from multi-component wave data. *J. Phys. Oceanogr*, **9**, 373.

**Lukacs, E.** (1970) : *Characteristic Functions*, 2nd edn., Griffin, London.

**Lygre, A.** and **Krogstad, H.E.** (1986) : Maximum entropy estimation of the directional distribution in ocean wave spectra. *J. Phys. Oceanogr.* **16**, 2052.

**Mardia, K.V.** (1972) : *Statistics of Directional Data*, Academic Press, London.

— (1975a) : Statistics of directional data (with discussions). *J. Roy. Statist. Soc. Ser. B*, **37**, 349.

— (1975b) : Characterizations of directional distributions. In G.P. Patil, S. Kotz & J.K. Ord (eds.), *Statistical Distributions in Scientific Work*, Reidel, Dordrecht, **3**, 365.

**Marshall, A.W.** and **Olkin, I.** (1979) : *Inequalities : Theory of Majorization and Its Applications.* Academic Press, New York.

**Ochoa, J.** and **Delgado-González. O.E.** (1990) : Pittalls in the estimation of wind wave directional spectra by variational principles, *Applied Ocean Research.* **12**, 180.

**Rattihalli, R.N.** and **SenGupta, A.** (2002) : Classes of asymmetric circular distributions *Tech. Rep.* 267, Dept. of Statistics, University of California, Riverside.

**Rukhin, A.L.** (1972) : Some statistical decisions about distributions on a circle for large sample, *Sankhyá*, Ser, A **34**, 243.

**SenGupta, A.** (1998) : *DDSTAP – Directional Data Statistical Analysis Package.* Indian Statistical Institute, Calcutta.

**Singh, H., Hnizdo, V.** and **Demchuk, E.** (2002) : Probabilistic model for two dependent circular variables. *Biometrika*, **89**, 719.

**Tucker, M.J.** (1991) : *Waves in Ocean Engineering-Measurement Analysis and Interpretation.* Ellis Holwood, New York.

**von Mises, R.** (1918) : Über die "Ganzzahligkeit" der Atomgewichte und Verwandte Fragen. *Physikal. Z.*, **19**, 490.

**Yfantis, E.A.** and **Borgman, L.E.** (1982) : An extension of the von Mises distribution. *Commun. Statist.*, **11**, 1695.

APPLIED STATISTICS UNIT
INDIAN STATISTICAL INSTITUTE
KOLKATA, INDIA
*E-MAIL* : ashis@isical.ac.in