A SIMPLE METHOD FOR ANALYZING THE TAILS OF DISTRIBUTIONS

por

Román Castañeda[1] & Juan Carlos Correa[2]

Resumen

Castañeda R. & J. C. Correa: A simple method for analyzing the tails of distributions. Rev. Acad. Colomb. Cienc. **23**(89): 555-562. 1999. ISSN 0370-3908.

Se sabe que sus momentos caracterizan a una distribución. Usando densidades generadas por la distribución que se estudia, podemos analizar partes de la distribución original que estén sombreadas por aquellas partes con altas densidades.

Palabras clave: Momentos, colas, densidad.

Abstract

Moments are known to characterize a distribution completely. Using densities that are generated from the distribution under study we can analyze partsof the original distribution that are shadowed by those parts with high densities.

Key words: Moments, heavy tails, density.

1. Intraduction

Moments we known to characterize a distribution completely. In practical situations we use at most moments up to order fourth. **Castañeda** (1993) uses a method to analyze characteristics of the tails of a symmetric distribution by using bigh-order moments of a bivariate distribution. The basic idea is to generate a family of distributions associated with the given distribution of the data. This family of associated distributions is generated in such a way that each member is related to the moments of the given distribution. Each of these distributions reveals characteristics in the original distribution by given more weight ta some parts of the distribution, for example, clusters or multimodality.

2. The method

Let us consider a continuaus diatribution with density $f(x)$ and moments with respect to the origin given by $\mu^{(k)} = E[X^k] = \int x^k f(x)dx$, $k = 1,2,\ldots$ If we are interested in analyzing some specific parts of the distribution, maybe those with low density that are shadowed by those parts with high density, then we would like to

1 Universidad Nacional deColombia, Sede Medellín. A.A. 3840, Medellín, Colombia e-mail: rcastane@peseus.unalmed.edu.co

2 Universidad Nacional deColombia, Sede Medellín. A.A. 3840, Medellín, Colombia.

give weights to $f(x)$ in a systematic way. From the moment definition the integrand is $x^{2k}f(x)$, that properly normalized is a new density function for $k = 1, 2, \cdots$. The polynomial x^{2k} gives different weight to different parts of the support of the distribution. The advantage of normalizing is that the moments of these new densities are related to those of the original distribution in a simple way:

$$f_k(x) = \frac{x^{2k} f(x)}{\mu_{2k}}, \ k = 1, 2, \cdots . \tag{1}$$

The random variable defined with this density has first and second moments that are related to the moments of the original distribution in the following way

$$\mu_1^{(k)} = \frac{\mu_{2k+1}}{\mu_{2k}} \text{ and } \mu_2^{(k)} = \frac{\mu_{2k+2}}{\mu_{2k}} . \tag{2}$$

If we consider $\mu_1^{(1)}$, the center of the first moment-generated distribution, and we compare this moment to the first moment of the original distribution we can build a measure of displacement that will be a measure of skewness.

If we consider moments with respect to an arbitrary point, say a, we can screen some parts of the original distribution that could be producing parts with high densities in the moment-generated distributions. This happens when the original distribution is highly skewed. We can consider the moments with respect to the mode to shadow the area with the highest density quickly.

$$f_k(x) = \frac{(x-a)^{2k} f(x)}{\mu_{2k}}, \ k = 1, 2, \cdots . \tag{3}$$

There is a problem when we use this procedure with highly skew distributions. The area with high density does not disappear and this does not permit us to analyze the tail where this concentration occurs. To avoid this problem a symmetrization is recommended. We can transform the data using a Box-Cox transformation.

In practice we do not know the true distribution but just a sample from that population. In that case we use a smooth estimate of the density, for example a kernel density estimate, say $\hat{f}(x)$ (**Silverman**, 1986). Then the associated distributions will be

$$f_k(x) = \frac{x^{2k} \hat{f}(x)}{\mu_{2k}}, \ k = 1, 2, \cdots . \tag{4}$$

It is well known that if the continuous random variable X has cumulative distribution F, then $Y = F(X)$ has a uniform distribution between 0 and 1. Thus the transformation $W = 2Y - 1$ has uniform distribution between -1 and 1. If we consider the polynomials x^k, with $k = 1, 2, \cdots$, we have that the sequence of polynomials goes to zero everywhere except at -1 and 1 as $k \to \infty$ in the interval $[-1, 1]$. Thus they give weights to the tails in a systematic way. The central part of the support receives less weight when the grade of the polynomial increases. In a practical situation what we do is the following: We have our sample $x_1, x_2, \cdots, x_n$. Let $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ be the order statistics. Then define

$$y_{(i)} = 2\left(\frac{x_{(i)} - x_{(1)}}{x_{(n)}}\right) - 1. \tag{5}$$

Then the data will be in the range -1 to 1.

3. The bivariate case

Let $f(x, y)$ be a two-dimensional pdf with support given by R. The (m, n)-moment will be defined by

$$M_{mn} = \int_R \int x^m y^n f(x, y)\, dx\, dy \tag{6}$$

m and n are non-negative integers. The sum $m + n$ determines the order of the moment. There are $m + n + 1$ moments of order $m + n$. It is well known that the first order moments M_{01} and M_{10} specify the position of the center of mass of the distribution. Let us assume that the condition $M_{10} = M_{01} = 0$ holds, i.e. the origin of the coordinates will be attached at the center of mass of the distribution. In practical situations R is a rectangular area, say $4AB$, centered at the center of mass of the distribution. We introduce the reduced coordinates

$$z = \frac{x}{A}, \ z = \frac{x}{A} \tag{7}$$

to scale the region R onto the interval $(-1, 1) \times (-1, 1)$ and define the centered reduced moments of the distribution by

(8)

$$\mu_{mn} = \frac{M_{mn}}{A^{m+1} + B^{n+1}} = \int_{-1}^{1} \int_{-1}^{1} z^m w^n p(z, w)\, dz\, dw .$$

The reduced centered moments are dimensionless and they become comparable with each other. Note that $|z|^m \geq |z|^{m+1}$, $|w|^n \geq |w|^{n+1}$ in $(-1, 1) \times (-1, 1)$. As a consequence the polynomials z^m and w^n will enhance certain portions of the distribution depending on m and n.

Let us define the associated density functions by

$$V_k(z, w) = \frac{\left(z^2 + w^2\right)^k}{\sum_{j=0}^{k} \mu_{2(k-j),2j}} p(z, w) , \tag{9}$$

where $k = 1, 2, \cdots$ denotes the order of the function and $z^2 + w^2 = r^2$ is the squared distance between the center of mass of the distribution and an arbitrary point

(z, w) in its domain. Note that $V_0(z,w) = p(z,w)$ but for $k > 0$ the coefficient of $V_k(z,w)$ weights the values of the distribution radially, in such a way that the values of the distribution around the center will be nullified whereas the values far form it will be enhanced. Thus, each associated density function refers to a specific zone of the original distribution. It allows to perform a precise evaluation of the distribution by segmentation.

The geometrical parameters of each associated pdf (i.e position of the center of mass, length and orientation of semiaxis, eccentricity) are used as quantitative descriptors of the structure of the respective zone. The moments, denoted by $\mu_{mn}^{(k)}$, are determined through the following combinations of the centered reduced moments of the original distribution,

$$\mu_{01}^{(k)} = \frac{\sum_{j=0}^{k}\binom{k}{j}\mu_{2(k-j),2j+1}}{\sum_{j=0}^{k}\binom{k}{j}\mu_{2(k-j),2j}} \tag{10}$$

$$\mu_{10}^{(k)} = \frac{\sum_{j=0}^{k}\binom{k}{j}\mu_{1+2(k-j),2j}}{\sum_{j=0}^{k}\binom{k}{j}\mu_{2(k-j),2j}} \tag{11}$$

$$\mu_{02}^{(k)} = \frac{\sum_{j=0}^{k}\binom{k}{j}\mu_{2(k-j),2(j+1)}}{\sum_{j=0}^{k}\binom{k}{j}\mu_{2(k-j),2j}} \tag{12}$$

$$\mu_{20}^{(k)} = \frac{\sum_{j=0}^{k}\binom{k}{j}\mu_{2(1+k-j),2j}}{\sum_{j=0}^{k}\binom{k}{j}\mu_{2(k-j),2j}} \tag{13}$$

$$\mu_{11}^{(k)} = \frac{\sum_{j=0}^{k}\binom{k}{j}\mu_{2(k-j)+1,2j+1}}{\sum_{j=0}^{k}\binom{k}{j}\mu_{2(k-j),2j}} \tag{14}$$

$\mu_{01}^{(k)}$ and $\mu_{10}^{(k)}$ specify the position of the center of mass of $V_k(z,w)$ respective to the center of mass of the original distribution. However, to determine the other geometrical parameters of $V_k(z,w)$ we define a new coordinate system with the origin at its center of mass and the axis parallel to its semiaxis respectively, i.e. the proper coordinate system of $V_k(z,w)$. First we calculate the moments with respect to the proper coordinate origin by means of the formula

$$\eta_{01}^{(k)} = \sum_{l=0}^{m}\sum_{j=0}^{n}(-1)^{k+j}\binom{m}{l}\binom{n}{j} z_{k0}^{l} w_{k0}^{j} \mu_{ml,n-j}^{(k)} . \tag{15}$$

So we obtain

$$\begin{bmatrix} \eta_{01}^{(k)} \\ \eta_{10}^{(k)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} , \tag{16}$$

$$\begin{bmatrix} \eta_{02}^{(k)} \\ \eta_{11}^{(k)} \\ \eta_{20}^{(k)} \end{bmatrix} = \begin{bmatrix} \mu_{02}^{(k)} \\ \mu_{11}^{(k)} \\ \mu_{20}^{(k)} \end{bmatrix} - \begin{bmatrix} \left(\mu_{01}^{(k)}\right)^2 \\ \mu_{01}^{(k)}\mu_{10}^{(k)} \\ \left(\mu_{10}^{(k)}\right)^2 \end{bmatrix} \tag{17}$$

Note that $\eta_{nm}^{(0)} = \mu_{nm}^{(0)} = \mu_{nm}$, with $n + m = 1,\ 2$. Now, the new second order moments are calculated in the proper coordinate system, whose axis are rotated an angle

$$\varphi_k = \frac{1}{2}\arctan\left(\frac{2\eta_{11}^{(k)}}{\eta_{20}^{(k)} - \eta_{02}^{(k)}}\right) \tag{18}$$

with respect to the original ones. There is an ambiguity in the tilt angle φ_k which could be solved in the same way as **Teague** (1980), by choosing φ_k always to be the angle between the x-axis and the semimajor axis a_k, i.e. $a_k \geq b_k$, and taking into account that

$$-\frac{\pi}{2} \leq \arctan x \leq \frac{\pi}{2},\ x = \frac{2\eta_{11}^{(k)}}{\eta_{20}^{(k)} - \eta_{02}^{(k)}} \tag{19}$$

With these conventions we arrive at the results for the tilt angle that are given in Table I.

TABLE I

$\eta_{20}^{(k)} - \eta_{02}^{(k)}$	$\eta_{11}^{(k)}$	φ_k	
Zero	Zero	0^0	
Zero	Positive	$+45^0$	
Zero	Negative	-45^0	
Positive	Zero	0^0	
Negative	Zero	-90^0	
Positive	Positive	$\frac{1}{2}\arctan x$	$0^0 < \varphi_k < 45^0$
Positive	Negative	$\frac{1}{2}\arctan x$	$-45^0 < \varphi_k < 0^0$
Negative	Positive	$\frac{1}{2}\arctan x + 90^0$	$45^0 < \varphi_k < 90^0$
Negative	Negative	$\frac{1}{2}\arctan x + 90^0$	$-90^0 < \varphi_k < -45^0$

Introducing the usual rotation of coordinates by φ_k we obtain

$$\begin{bmatrix} \eta_{02}^{\prime(k)} \\ 0 \\ \eta_{20}^{\prime(k)} \end{bmatrix} = \begin{bmatrix} \cos^2\varphi_k & 2\cos\varphi_k\sin\varphi_k & \sin^2\varphi_k \\ -\cos\varphi_k\sin\varphi_k & \cos^2\varphi_k & \cos\varphi_k\sin\varphi_k \\ \sin^2\varphi_k & -2\cos\varphi_k\sin\varphi_k & \cos^2\varphi_k \end{bmatrix} \begin{bmatrix} \left(\mu_{01}^{(k)}\right)^2 \\ \eta_{02}^{(k)}\eta_{11}^{(k)} \\ \left(\eta_{20}^{(k)}\right)^2 \end{bmatrix} \tag{20}$$

4. Examples

To illustrate the above ideas we present several examples. Let us consider the data about the cost of health services provided by the university health department to its employees. The distribution is highly skewed as is shown by its estimate in Figure 1a. The first moment distribution is shown in Figure 1b. The first-moment-with-respect-to-the-mode density estimate is shown in Figure 1c.

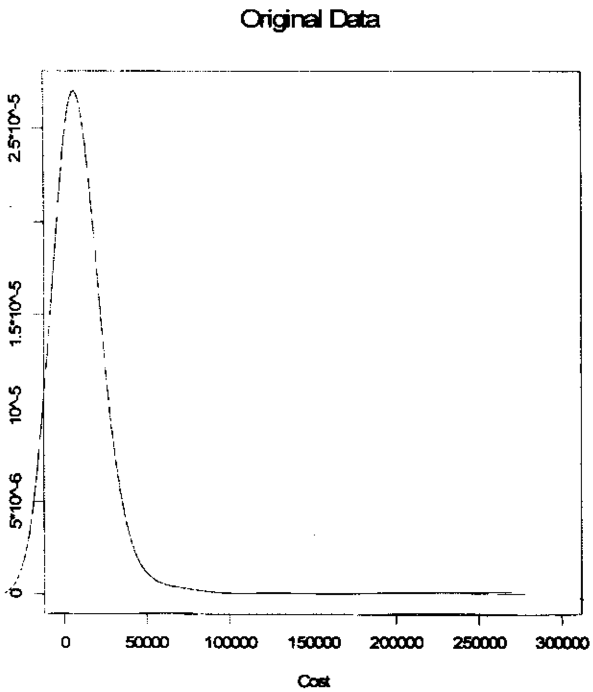

Figure 1 (a)

We can see how the high-density part has disappeared and others parts start appearing showing multimodality. This behavior has a logical explanation for this problem. The most of the services required by users at the health department are nonexpensive check-outs for well-known diseases such as colds, simple headaches, and so on, or small accidents. However a small fraction of the requirements are related with more expensive problems: heart problems, etc.

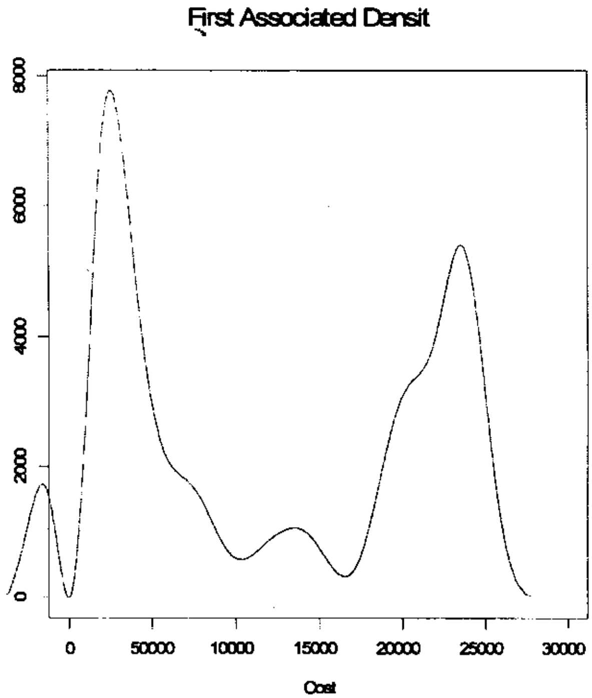

Figure 1 (b)

The next univariate example shows a normal distribution. The associated distributions are bimodal. The higher the grade of the associated distribution, the further apart the modes are. The relations between the main distribution and its associated distributions give us new measures of skewness, for example the difference between the means of the main and the first associated distributions (Figs. 2a-2d). In Figure 2b we observe a third mode on the right. The explanation for this mode is that the random generator mechanism for the normal distribution has problems generating values from the tails. This is a common problem with random number generators.

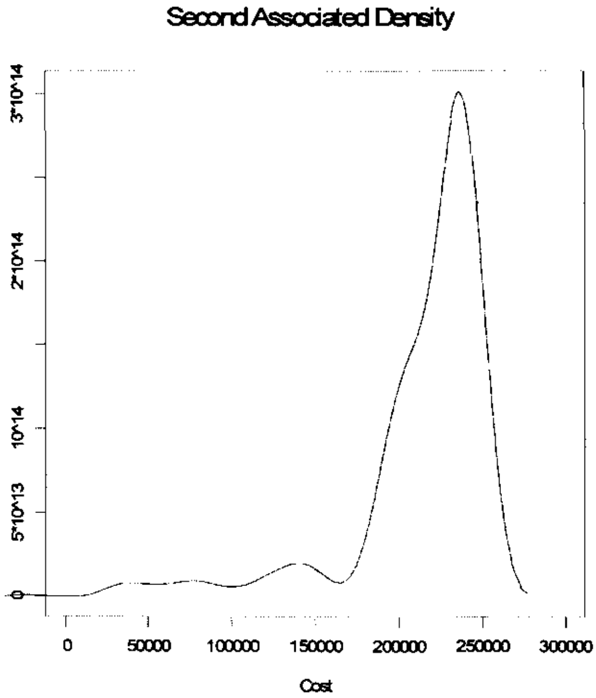

Figure 1 (c)

The next example is important in optics. It is well known that PSFs (Point Spread Functions) have complete information of the imaging properties of optical systems. They are defined as the intensity distribution in the image plane of the system when the object is deltalike (**Teague**, 1980). PSFs of diffraction limited systems are rotation symmetric and about 84% of the energy is concentrated in the Airy disc. System aberrations change this symmetry and/or this energy distribution. Therefore a precise evaluation of image quality can be obtained from the quantitative analysis of both the symmetry and the energy distribution of the PSFs. Associated density functions are introduced to perform the analysis of the central disc and the ring structure of the PSF separately. The geometrical parameters of those functions are determined from the centred reduced moments of the PSF and are used as symmetry descriptors. Thus, aberration types can be characterized using descriptors, which are determined directly from experimental data.

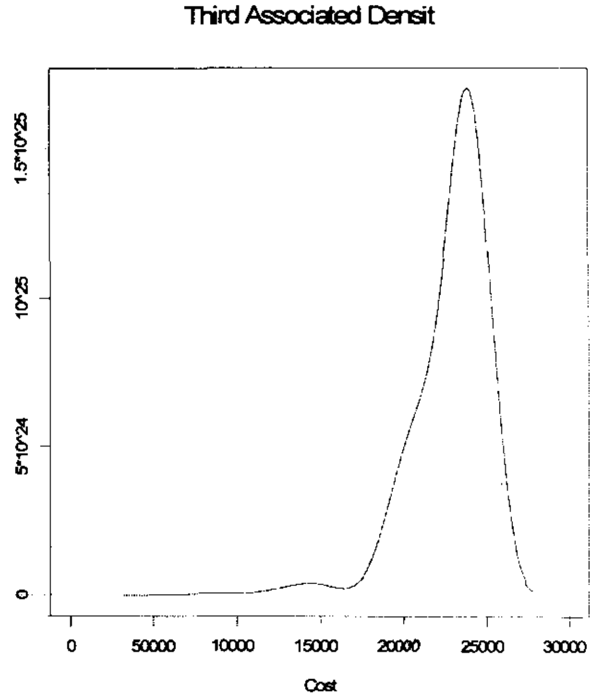

Figure 2 (a)

Measured PSFs of many systems of interest, such as microscope lenses, consist in an intensity peak surrounded by a ring structure. Up a finite distance from the peak their intensity values will be immersed into the background noise of the detector and can be neglected. In this work we only concern on this class of PSFs.

Let us assume that the energy content of the PSF into the region R arround its intensity peak determined by such a distance is equal one. Thus, the PSF can be interpreted as a real density function, whose symmetry and energy distribution can be described using their moments (**Marathay**, 1982), which are defined as (6).

We will introduce reduced coordinates to scale the region R onto the interval $(-1, 1)$ and to obtain dimensionless centred reduced moments (8).Furthermore, to evaluate the symmetry and the energy content of the PSF it is useful to introduce associated density functions (9). If the system is diffraction limited, the associated density functions will determine two concentric isotropic zones. Figs. 3a-3b show these zones for the on-axis PSF of a microscope lens $60\times/0.80$ at wavelength of 589nm.

But aberrations introduce geometrical variations of the zones. Fig. 5 shows the zones for an off-axis PSF of a microscope lens $40 \times /0.65$ at wavelength of 486nm. The geometrical parameters of the associated density functions (i.e. position of centers of mass, length and orientation of the semi-axis, eccentricity) will be used as quantitative descriptors of the symmetry and energy distribution of the zones.

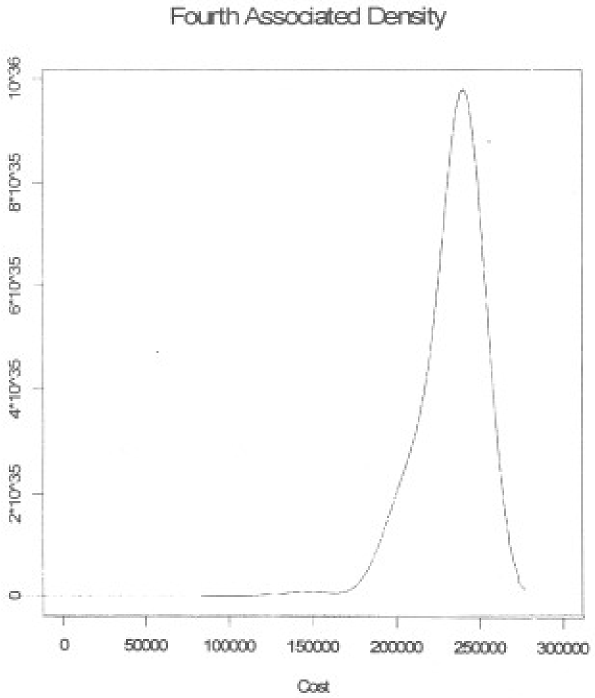

Figure 2 (b)

If the system is diffraction limited, the associated density functions will determine two concentric isotropic zones. Figs. 3a-3b show these zones for the on-axis PSF of a microscope lens $60 \times /0.80$ at wavelength of 589nm.

But aberrations introduce geometrical variations of the zones. Fig. 5 shows the zones for an off-axis PSF of a microscope lens $40 \times /0.65$ at wavelength of 486nm. The geometrical parameters of the associated density functions (i.e. position of centers of mass, length and orientation of the semi-axis, eccentricity) will be used as quantitative descriptors of the symmetry and energy distribution of the zones.

Using V_0 we analyze the central disc of the PSF principally, i.e. a zone that includes the center of mass of the PSF and its surroundings but does not coincide with the Airy disc of the PSF in general (Figs. 3a and 4a inside the rectangle). Its geometrical parameters are determined by the centred reduced moments up to the second order.

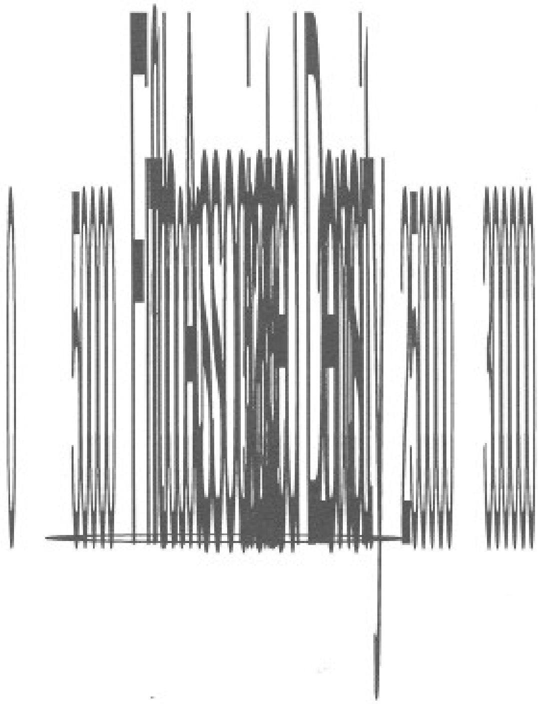

Figure 2 (c)

The semiaxis of V_0 are oriented at an angle φ_0 (18). Introducing the usual rotation of coordinates by this angle for $m + n = 2$ in eq.(20), we obtain new centred reduced moments of second order. The length of the semiaxis and the eccentricity of V_0 can be expressed by

$$(21) \qquad a_0 = \sqrt{\frac{\mu'_{20}}{\mu'_{00}}}\,, b_0 = \sqrt{\frac{\mu'_{02}}{\mu'_{00}}}$$

$$(22) \qquad E_0 = \sqrt{1 - \left(\frac{b_0}{a_0}\right)^2} = \sqrt{1 - \frac{\mu'_{02}}{\mu'_{00}}}$$

respectively. We assume that $a_0 \geq b_0$, so that $0 \leq E_0 < 1$. For rotation symmetrical V_0 we have $\varphi_0 = 0, E_0 = 0, a_0 = b_0 = R_0$. R_0 denotes the reduced radius of the central disc of the PSF, which takes its minimum value for diffraction limited systems.

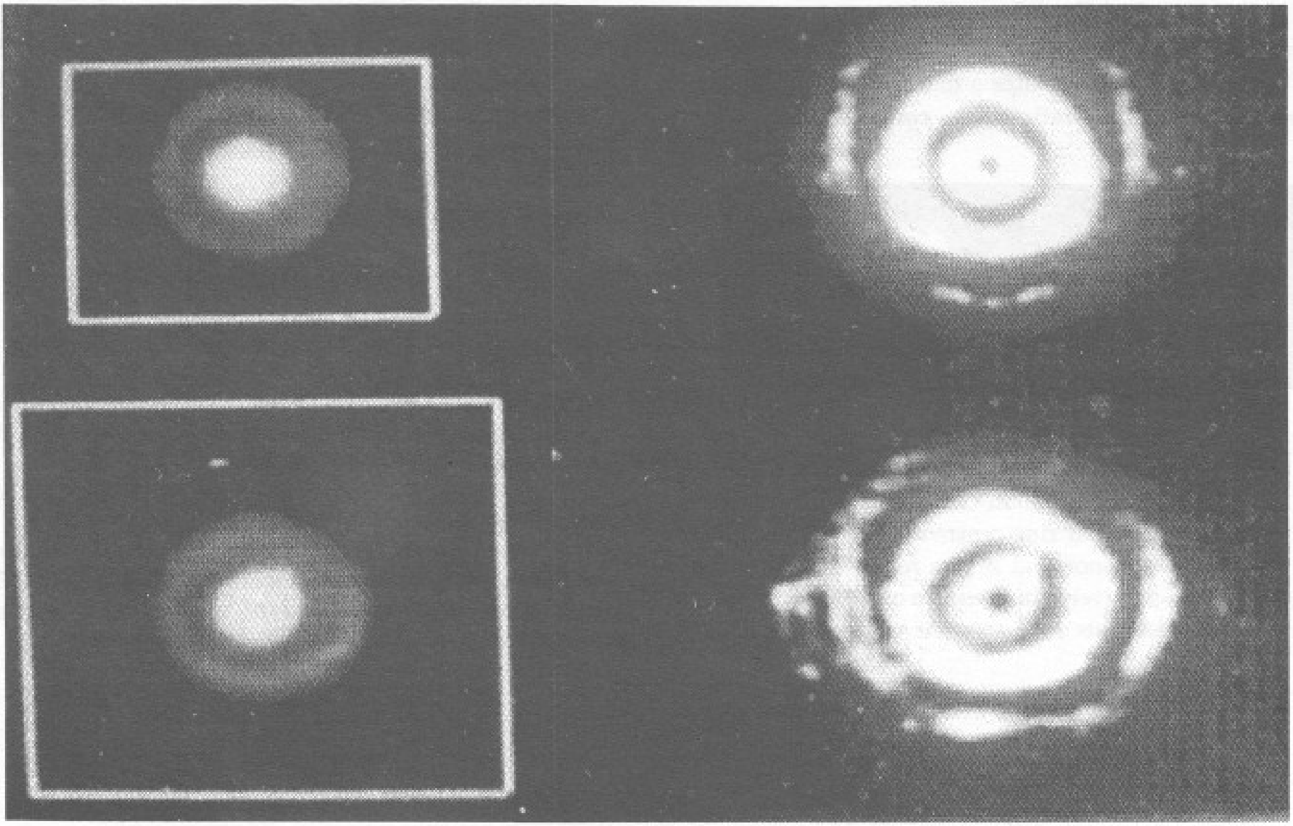

Figure 3

Rotation symmetrical aberrations (i.e. spherical aberration, defocus, cuvature of field) increase the value of R_0 but do not change the symmetry of V_0. In presence of non rotation symmetrical aberrations (i.e. astigmatism, coma) the central disc will be not rotation symmetric but its mean-square radius will be

$$\epsilon_0^2 = \frac{1}{2}\left(a_0^2 + b_0^2\right) \quad (23)$$

If we regard $b_0 \geq 0.9a_0$ as permissible for rotation symmetry, we obtain $E_0 \geq 0.44$ as a tolerance.

Using V_1 we concern with the zone that excludes the center of mass of the PSF and its neighborhood and includes the ring structure of the PSF near the central disc (Figs. 3b and 4b right). Its geometrical parameters will be determined by the centred reduced moments of third and fourth orders. The semiaxis of are oriented at an angle φ_1, which is given by (18) for $k = 1$.

Let us introduce the rotation of coordinates by for $m + n = 3, 4$ in eq. (8) to obtain new centred reduced moments of third and fourth orders (with primes). As expected, we obtain $\mu'_{13} = \mu'_{31} = 0$. The coordinates (Z_1, W_1) of the center of mass, the length of the semiaxis and the eccentricity are given by

$$Z_1 = \frac{\mu'_{30} + \mu'_{12}}{\mu'_{02} + \mu'_{20}}, \quad W_1 = \frac{\mu'_{03} + \mu'_{21}}{\mu'_{02} + \mu'_{20}}, \quad (24)$$

$$a_1 = \sqrt{\frac{\mu'_{40} + \mu'_{22}}{\mu'_{02} + \mu'_{20}}}, \quad a_1 = \sqrt{\frac{\mu'_{04} + \mu'_{22}}{\mu'_{02} + \mu'_{20}}}, \quad (25)$$

$$E_1 = \sqrt{1 - \left(\frac{b_1}{a_1}\right)^2} = \sqrt{1 - \frac{1 + \mu'_{04}/\mu'_{22}}{1 + \mu'_{40}/\mu'_{22}}} \quad (26)$$

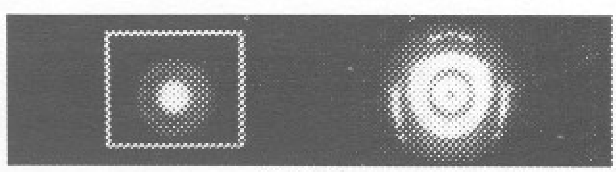

Figure 4

Centers of mass, lengths of semiaxis and eccentricities are rotation invariants. Thus, the geometrical parameters of and are independent in the sense that they are functions of centred reduced moments of different orders.

The vector $\vec{Q} = \vec{Z}_1 + \vec{W}_1$ represents the shift of the center of mass of V_1 with respect to the center of mass of V_0 and thus, it constitutes a descriptor for coma. Indeed, we will call the ratio $|\vec{Q}|/\epsilon_0$ coma factor.

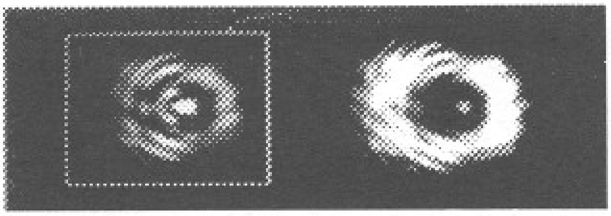

Figure 5

If V_1 is rotation symmetric we have $\varphi = 0$, $\vec{Q}_1 = 0$, $a_1 = b_1 = R_1$, $E_1 = 0$. In this case the zone relative to V_1 will be a circular ring centred in the center of mass of the PSF. Its minor and major reduced radii will be R_0 and R_1 respectively. In presence of rotation symmetrical aberrations this zone remains concentric with the central disc but the value of R_1 will increase. For non rotation symmetrical aberrations the geometry of the PSF can be very complicated because the centers of mass, the eccentricities and the axis orientation of V_0 and V_1 can be different. However, if we regard $|\vec{Q}|/\epsilon_0 \leq 0.2$ as permissible for coma free systems, we can define a mean circular ring of minor radius ϵ_0 and major radius

$$\epsilon_1 = \sqrt{\frac{1}{2}\left(a_0^2 + b_0^2\right)} .$$

References

[1] **Castañeda, R.**, *Two-Dimensional Evaluation of Point Spread Functions Using Centred Reduced Moments*, Pure Appl. Opt. **3** (1994), 737-755.

[2] **Marathay, A.**, *Elements of optical coherence theory*, John Wiley & Sons, New York, 1982, p. 232.

[3] **Silverman, B. W.**, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986, 1980.

[4] **Teague, M.**, *Image analysis via the general theory of moments*, J. Opt. Soc. Am. **70** (1980), 920-930.