DOI: 10.13203/j.whugis1997.01.012

第 22卷 第 1期 武汉测绘科技大学学报 Vol. 22 No. 1
1997年 3月 Journal of Wuhan Technical University of Surveying and Mapping March 1997

地图手工数字化对点误差的统计分布

史文中 刘文宝

(香港理工大学土地测量及地理资讯学系,香港九龙红磡)

摘 要 定义了对点误差,导出了统计分布密度和分布函数,给出了矩母函数、特征函数和数字特征,以及对点误差及其投影分量和标准化分布。

关键词 数字化;对点误差;投影分量;概率分布

分类号 P207, P289

当前,GIS的主要数据来源之一仍然是现有地图的手工数字化。为了有效地控制数字化点的位置精度,人们历来十分重视研究离散点的数字化误差。但以往的研究大多局限于实验方法,例如文献[1~6],所得结论仅是数值或定性的,无法揭示对点误差的本质。本文从理论上探讨对点误差的统计分布及其数字特征,以便为生产中正确分析手工数字化误差奠定基础。

1 对点误差及其投影分量的定义

图 1中 $P'(x',y')$ 和 $P(x,y)$ 分别为某点

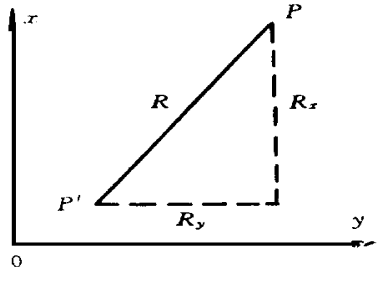


图 1 对点误差

Fig. 1 Centering Error

在地图上的位置及手工数字化后的位置。将 P 与 P' 点间的距离 $R=|PP'|$ 定义为手工数字化对点误差,它在 x、y 轴方向上的投影记为 R_x和 R_y。显然,有:

$$\begin{cases} R_x = x - x', \quad R_y = y - y' \\ R = \sqrt{R_x^2 + R_y^2} \end{cases} \tag{1}$$

由于 x 和 y 均为随机变量,因而 R_x、R_y和 R 也将是随机变量。Bolstad等人(1990)[6]曾用实验方法分析过 R_x和 R_y的经验分布。尔后,Caspary & Scheuring (1993)[7]又从理论上进一步探讨了 R_x和 R_y的标准化分布,而 R 的概率分布将是本文讨论的重点之一。

2 对点误差的分布函数和分布密度

由于对点误差 R 的投影分量 R_x和 R_y服从正态分布[7],在 R_x和 R_y相互独立且方差相等时,R_x和 R_y的联合分布密度为:

$$f_{R_x,R_y}(r_x,r_y) = \exp\{-[(r_x-\mu_x)^2 + (r_y-\mu_y)^2]/(2\sigma^2)\}/(2\pi\sigma^2) \tag{2}$$

其中 σ^2为 R_x和 R_y的方差。根据随机变量分布函数的定义知[8],对点误差 R 的分布函数为:

$$F_R(r) = \iint_K \frac{1}{2\pi\sigma^2} \exp\{[-\frac{1}{2\sigma^2}][(r_x-\mu_x)^2 + (r_y-\mu_y)^2]\}\, dr_x\, dr_y \tag{3}$$

其中积分区域 K为以 $P'(\mu_x,\mu_y)$ 为中心、r 为半径的圆盘,即 $(r_x^2+r_y^2)^{1/2}\leqslant r$。(3)式经变量代换后,再对 r 求导数,略去中间推导[9],得对点误差 R 的分布密度:

$$f_R(r) = \begin{cases} \frac{r}{\sigma^2}\exp\left\{-\frac{r^2+\mu^2}{2\sigma^2}\right\} I_0\left(\frac{r\mu}{\sigma^2}\right) \\ \qquad\qquad \text{当 } r \geqslant 0\text{时} \\ 0, \qquad\qquad \text{当 } r < 0\text{时} \end{cases} \tag{4}$$

而相应的数学期望和方差分别为:

$$\mu_R = \sigma\sqrt{\frac{\pi}{2}}\left[(1+\frac{\mu^2}{2\sigma^2})I_0(\frac{\mu^2}{4\sigma^2}) + \frac{\mu^2}{2\sigma^2}I_1(\frac{\mu^2}{4\sigma^2})\right]\exp(-\frac{\mu^2}{4\sigma^2}) \tag{5}$$

$$\sigma_R^2 = 2\sigma^2 + \mu^2 - \mu_R^2 \tag{6}$$

收稿日期:1996-10-08。史文中,男,33岁,助理教授,博士生导师,现从事 RS& GIS集成中不确定性研究。

其中 $\mu = \sqrt{\mu_x^2 + \mu_y^2}$ 为对点误差中的系统误差部分，$I_0(\cdot)$ 和 $I_1(\cdot)$ 为实变量的第二类 Bessel 函数[10]。当 $\mu = 0$时，(4) 式简化为：

$$f(r) = \begin{cases} \dfrac{r}{\sigma^2} \exp\left\{-\dfrac{r^2}{2\sigma^2}\right\}, & \text{当 } r \geqslant 0\text{时} \\ 0, & \text{当 } r < 0\text{时} \end{cases} \tag{7}$$

(7) 式即为 Rayleigh 分布密度函数。这说明在只有随机误差成份时，对点误差 R 是一个服从 Rayleigh 分布的随机变量，其分布密度如图 2

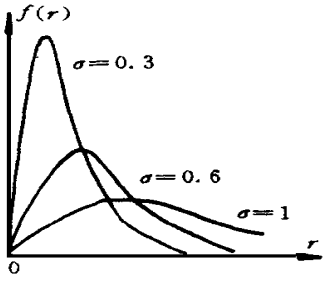


图 2 对点误差的分布密度

Fig. 2 Centering Error's Density Distribution

利用 (7) 式可得对点误差 R 的分布函数：

$$F(r) = \begin{cases} 1 - \exp\left\{-\dfrac{r^2}{2\sigma^2}\right\}, & \text{当 } r \geqslant 0\text{时} \\ 0, & \text{当 } r < 0\text{时} \end{cases} \tag{8}$$

3 对点误差分布的矩母函数和特征函数

3.1 矩母函数

根据随机变量矩母函数的定义[8]：

$$J_r(t) = E(e^{tr}) \tag{9}$$

又因对点误差 R 为连续型随机变量，故有：

$$J_r(t) = \int_{-\infty}^{+\infty} e^{tr} f(r)\, dr \tag{10}$$

将 (7) 式代入 (10) 式后得：

$$J_r(t) = \exp\left\{\frac{\sigma^2 t^2}{2}\right\} \int_0^{+\infty} \frac{r}{\sigma^2} \cdot \exp\{-(r - \sigma^2 t)^2/(2\sigma^2)\}\, dr$$

对上式作变量代换 $(r - \sigma^2 t)/\sigma = z$，有：

$$J_r(t) = 1 + \exp\{\sigma^2 t^2/2\}\sigma t \cdot \left[\int_0^{+\infty} \exp\left\{-\frac{z^2}{2}\right\} dz + \int_{-\sigma t}^{0} \exp\left\{-\frac{z^2}{2}\right\} dz\right] \tag{11}$$

顾及 $\int_0^{+\infty} \exp\left\{-\dfrac{z^2}{2}\right\} dz = \sqrt{\dfrac{\pi}{2}}$，并将 $\exp\{-z^2/2\}$ 按幂级数展开，代入 (11) 式得：

$$J_r(t) = 1 + \exp\{\sigma^2 t^2/2\}\sigma t\left[\sqrt{\pi/2} + \sigma t - (\sigma t)^3/6 + (\sigma t)^5/40 - (\sigma t)^7/336 + \cdots - (-1)^{n+1}(\sigma t)^{2n+1}/(2^n n!(2n+1)) + \cdots\right] \tag{12}$$

上式即为对点误差 R 的矩母函数。

3.2 特征函数

仿证明矩母函数的方法，可得对点误差 R 的特征函数为：

$$H(t) = 1 + \exp\{(i\sigma t)^2/2\} i\sigma t\left[\sqrt{\pi/2} + i\sigma t - (i\sigma t)^3/6 + (i\sigma t)^5/40 - (i\sigma t)^7/336 + \cdots - (-1)^{n+1}(i\sigma t)^{2n+1}/(2^n n!(2n+1)) + \cdots\right] \tag{13}$$

4 对点误差统计分布的数字特征

4.1 矩

将 (12) 式的矩母函数对 t 求各阶导数，令 $t = 0$后可得下列 4阶以下的原点矩：

$$\begin{cases} \mu'_1 = J'_r(0) = \sigma\sqrt{\pi/2} \\ \mu'_2 = J''_r(0) = 2\sigma^2 \\ \mu'_3 = J'''_r(0) = 3\sigma^3\sqrt{\pi/2} \\ \mu'_4 = J^{(4)}_r(0) = 8\sigma^4 \end{cases} \tag{14}$$

根据中心矩和原点矩之间的关系[8]，又可得相应的各阶中心矩：

$$\begin{cases} \mu_1 = 0 \\ \mu_2 = \mu'_2 - \mu'^2_1 = (2 - \pi/2)\sigma^2 \\ \mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1 \\ \quad = \sqrt{\pi/2}(\pi - 3)\sigma^3 \\ \mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1 \\ \quad = (8 - 3\pi^2/4)\sigma^4 \end{cases} \tag{15}$$

4.2 数学期望和方差

根据随机变量数学期望和方差的定义[8]，可得对点误差 R 的数学期望和方差：

$$\mu_R = E(r) = \mu'_1 = \sigma\sqrt{\pi/2} \tag{16}$$

$$\sigma_R^2 = D(R) = \mu_2 = (2 - \pi/2)\sigma^2 \tag{17}$$

应当注意，尽管对点误差 R 中的系统误差成份 $\mu = 0$，但由 (16) 式知 R 的数学期望并不为零，而取值 $\sqrt{\pi/2}\sigma$。这似乎说明具有最大概率的数字化点仍要偏离目标点。但事实上，数字化点的散布中心仍在目标点上。下面从理论上解释这一问题。为此，首先从另一种途径导出对点误差的分布密度。

引入下列极坐标 (r, θ) 变换：

$$\begin{cases} r_x = r\cos\theta \\ r_y = r\sin\theta \end{cases} \text{和} \begin{cases} \mu_x = \mu\cos\theta_0 \\ \mu_y = \mu\sin\theta_0 \end{cases} \tag{18}$$

则 (2) 式的指数部分变为:

$$[(r_x - \__x)^2 + (r_y - \__y)^2]/(2e^2) = [r^2 + \_^2 - 2r\_\cos(\theta - \theta_0)]/(2e^2) \quad (19)$$

由于 (18) 式的雅可比行列式:

$$J = \begin{vmatrix} \partial r_x/\partial r & \partial r_x/\partial\theta \\ \partial r_y/\partial r & \partial r_y/\partial\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r \quad (20)$$

故随机向量 (R,Θ) 的联合分布密度为:

$$g(r,\theta) = f(r\cos\theta, r\sin\theta)J = r\exp\{-1/(2e^2)[r^2 + \_^2 - 2r\_\cos(\theta - \theta_0)]\}/(2\pi e^2) \quad (21)$$

显然,对点误差 R 的分布密度就是 $g(r,\theta)$ 沿 r 的边缘分布,即有:

$$f(r) = \int_0^{2\pi} g(r,\theta)\,d\theta = (r/e^2)\exp\{-(r^2 + \_^2)/(2e^2)\}I_0(\_r/e^2) \quad (22)$$

这与 (4) 式相同。而 $g(r,\theta)$ 沿 θ 的边缘分布,即 $\Theta = \arctan(r_y/r_x)$ 服从 $[0, 2\pi]$ 内的均匀分布。因此,尽管对点误差的数学期望并不为零,但由于数字化点还沿 θ 呈均匀分布,则对点误差分布密度最大的数字化点将形成一个圆,圆心在目标点,半径为 $\sqrt{\pi/2}e$。这样,对点误差分布密度最大的数字化点的分布中心仍在目标点上。

4.3 众　数

在 (7) 式中,令 $df(r)/dr = 0$,可得对点误差 R 的众数:

$$M = e \quad (23)$$

这说明 R 的最大可能值等于坐标投影分量的标准差。又由 (16) 式知,$\__R = \sqrt{\pi/2}e \approx 1.25e$,因此,对点误差的众数位于数学期望之左,如图 3。

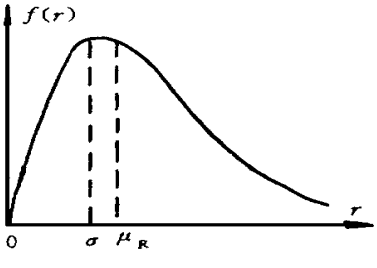


图 3　R 的众数和期望的相互位置

Fig. 3　Mutual Position of R's Mode and Expectation

4.4 偏态参数和峰态系数

根据定义[8],可得对点误差 R 的偏态系数 V_1 和峰态系数 V_2:

$$V_1 = \__3/[D(R)]^3 = [2(\pi - 3)/(4 - \pi)] \sqrt{\pi/(4-\pi)} \approx 0.62 \quad (24)$$

$$V_2 = \__4/[D(R)]^4 - 3 = (32 - 3\pi^2)/(4-\pi)^2 - 3 \approx 0.27 \quad (25)$$

由于 $V_1 > V_2 > 0$,因此,对点误差 R 的分布是正偏的,且较相应的正态分布尖瘦。

5　对点误差的标准化分布

对点误差投影分量 R_x 和 R_y 的方差样本估计式为[9]:

$$\begin{cases} \hat{e}_x^2 = [1/(n-1)]\sum_{i=1}^{n}(R_{x_i} - \hat{\_}_x)^2 \\ \hat{e}_y^2 = [1/(n-1)]\sum_{i=1}^{n}(R_{y_i} - \hat{\_}_y)^2 \\ \hat{e}_{xy} = [1/(n-1)]\sum_{i=1}^{n}(R_{x_i} - \hat{\_}_x)(R_{y_i} - \hat{\_}_y) \end{cases} \quad (26)$$

其中 $\hat{\_}_x$ 和 $\hat{\_}_y$ 分别为 R_x 和 R_y 的数学期望 $\__x$ 和 $\__y$ 的估值,估计公式为:

$$\hat{\_}_x = (1/n)\sum_{i=1}^{n} R_{x_i}, \quad \hat{\_}_y = (1/n)\sum_{i=1}^{n} R_{y_i} \quad (27)$$

利用 (21) 式构成点位协方差阵估计式:

$$\hat{D} = \begin{bmatrix} \hat{e}_x^2 & \hat{e}_{xy} \\ \hat{e}_{xy} & \hat{e}_y^2 \end{bmatrix} = \hat{D}^{1/2}(\hat{D}^{1/2})^{\mathrm{T}} \quad (28)$$

其中 $\hat{D}^{1/2}$ 为协方差阵 $\hat{D}$ 的 Cholesky 分解,且有:

$$\hat{D}^{1/2} = \begin{bmatrix} \sqrt{\hat{e}_x^2 - \hat{e}_{xy}^2/\hat{e}_y^2} & \hat{e}_{xy}/\hat{e}_y \\ 0 & \hat{e}_y \end{bmatrix} \quad (29)$$

于是,$\hat{D}^{1/2}$ 的逆阵为:

$$\hat{D}^{-1/2} = [\hat{D}^{1/2}]^{-1} = \frac{1}{\sqrt{\hat{e}_x^2\hat{e}_y^2 - \hat{e}_{xy}^2}} \begin{bmatrix} \hat{e}_y & -\hat{e}_{xy}/\hat{e}_y \\ 0 & \sqrt{\hat{e}_x^2 - \hat{e}_{xy}^2/\hat{e}_y^2} \end{bmatrix} \quad (30)$$

利用 (24) 式将由 (1) 式求得的对点误差投影分量向量标准化,得到一个服从标准化正态分布的独立误差向量 Z:

$$Z = \begin{pmatrix} Z_x \\ Z_y \end{pmatrix} = \frac{1}{\hat{D}^{1/2}} \begin{pmatrix} R_x - \hat{\_}_x \\ R_y - \hat{\_}_y \end{pmatrix} \sim N(0,1) \quad (31)$$

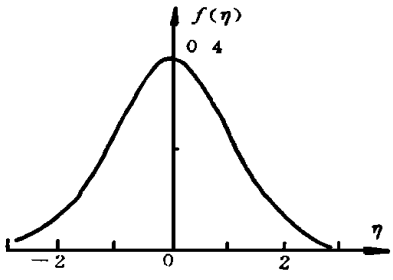
根据 χ^2 分布的定义知,标准化的对点误差平方量 Z_R^2 服从 χ^2 分布:

$$Z_R^2 = Z_x^2 + Z_y^2 = Z^{\mathrm{T}}Z \sim \chi_f^2 \quad (32)$$

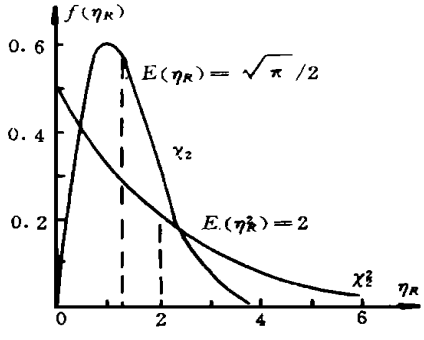
其中 $f = 2, E(Z_R^2) = 2, D(Z_R^2) = 4$。而标准化的对点误差 Z_R 服从 χ_2 分布:

$$Z_R = \sqrt{Z_x^2 + Z_y^2} \sim \chi_2 \quad (33)$$

其中 $E(Z_R) = \sqrt{\pi/2}, D(Z_R) = 2 - \pi/2$,这是 (16)、(17) 式中 $e = 1$ 时的特例。上述 (31)、(32) 和 (33) 式 3种分布密度函数的图像如图 4。



(a) $N(0,1)$ 正态分布



(b) i^2分布和 i_2分布

图 4 分布密度函数图像

Fig. 4 Function Image of Distribution Density

6 结 论

a. 当数字化数据中不含系统误差时,对点误差服从 Rayleigh 分布. 否则,服从分布密度为(4)式的广义分布.

b. 当数字化数据中不含系统误差时, $E(d)=0$,这里 $d=(R_x, R_y)^T$,有 $Z_R^2 = d^T D^{-1} d$服从中心 i^2分布,而 Z_R服从 i_2分布. 否则,将分别服从相应的非中心分布.

以上两点可用于对手工数字化数据中是否含有系统误差进行统计检验.

参 考 文 献

1 Baugh I D H, Borcham J R. Measuring the Eoastline from Maps- e Study of the Scottish Mainland. The Cartographic Journal, 1976, 13(2): 167~171

2 Traylor C. The Evaluation of a Methodology to Measure Manual Digitizing Error in Cartographic Databases. University of Kansas, 1979.

3 Burrough P A. Principles of Geographical Information Systems for Land Resources Assessment. Oxford: Clarendon, 1986. 116~118

4 Maffini G, Arno M, Bitterlich W. Observations and Comments on the Generation and Treatment of Error in Digital GIS Data. In: Accuracy of Spatial Databases. New York: Taylor and Francis, 1989. 55~67

5 Dunn R, Harrison A R, White J C. Positional Accuracy and Measurement Error in Digital Databases of Land Use: an Empirial Study. Int J GISs, 1990, 4(4): 385~398

6 Bolstad P V, Gessler P, Lillesand T M. Positional Uncertainty in Manually Digitized Map Data. Int J GISs, 1990, 4(4): 131~139

7 Caspary W, Scheuring R. Positional Accuracy in Spatial Databases. Comput., Enviror. and Urban Systems, 1993, 17(2): 103~110

8 李庆海,陶本藻. 概率统计原理和在测量中的应用. 北京:测绘出版社, 1982.

9 刘文宝. GIS空间数据的不确定性理论: [学位论文] 武汉:武汉测绘科技大学, 1995

10 《数学手册》编写组. 数学手册. 北京:高等教育出版社, 1979.

The Statistical Distribution of Cross-point Centering Error on Manual Digitizing a Map

Shi Wenzhong *Liu Wenbao*

(Department of Land Surveying & Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong)

Abstract The mouse centering error associated with manual digitizing a map was defined in this paper. The statistical distribution density and distribution function were driven. Moreover, moment generating function, characteristic function, digital characters, standard distribution and projection element of the error were discussed.

Key words digitizing; centering error; projection element; probability distribution