

# 无机岩石中有机质丰度的概率分布

杨文宽

(地质矿产部石油地质综合大队)

## 一、问题的提出

表征沉积岩所含有有机质<sup>1)</sup>丰度的数据,特别是有机碳丰度实测数据和氯仿沥青丰度实测数据,是石油地质工作者经常用到的数据。目前人们对这些数据的统计和应用,主要采用个别值法、平均值法、数值区间法和正态分布法,这四种方法在不同程度上均不甚完善,现简述如下。

1.个别值法 把每一块岩石样品的有机碳丰度实测数据都看成是十分准确的、代表性很强的数据,从而根据这些独立的彼此无关的数据在地层柱状图上划出达到“生油岩有机碳标准”的“生油层”和达不到标准的“非生油层”,这就是采用个别值法的实例。在实际工作中运用这种方法时,如果采样剖面的岩石成分相当单调(例如自底至顶,全部是块状灰岩),就很难判断某一个实测值所代表的真厚度究竟是多少米,很难把生油层与相邻非生油层的具体界面定下来,即使缩短采样间距也不能很好地解决这个划界问题。因为经验表明,有机碳在岩石中的分布很不均匀,即使是从同一块岩石样品上敲下来的几个碎块,它们各自的有机碳丰度一般也

互不相同,甚至出现有的碎块“达标”而另一些碎块“不达标”的情况。

2.平均值法 平均值的代表性无疑高于个别值,但是平均值法也并不完善。例如,采自湖南武岗县晏田剖面石磴子组的13块岩石样品(该组岩石成分单调,样品均为灰岩,其中12块样品的方解石含量 $\geq 95\%$ ),有机碳丰度平均值仅约0.13%,如果我们把“生油岩有机碳标准”定为0.15%,并且用这个标准去衡量平均值,那么我们会错误地认为晏田一带的石磴子组全部岩石都达不到标准。但是,事实上,这13块样品中,有机碳丰度 $\geq 0.20\%$ 者占23%, $\geq 0.30\%$ 者占8%,即使把标准再提高一些,仍有一部分岩石达到标准。

3.数值区间法 所谓数值区间法,就是用一个数值区间来表达某一个地层间隔(层、段、或组)的有机质丰度。例如用“0.04%至0.31%”来表达上述晏田石磴子组的有机碳丰度。但是,由于样品是随机采集的,0.04%并不一定是客观的最小值。如果再多采一些样品,最小值和最大值都可能改变,极差(区间长度)也可能随之而变。

1)本文所说的“有机质”,泛指以碳为主要成分且起源于生物的物质。

4. 正态分布法 在承认无机岩石中的有机质分布极不均匀和实验工作存在难于避免的技术性误差的基础上, 以有机碳或氯仿沥青丰度实测数据作为随机变量, 并且用数理统计的方法处理这些数据, 这是正态分布法优于以上三种方法的地方。但是, 关于正态分布(亦称常态分布)的假设, 实际上是一种先验性猜想。经验表明, 某些有机岩(例如油页岩)的有机质丰度, 可能大体上符合正态分布, 但是无机岩的有机质丰度, 一般并不符合正态分布。

根据以上的讨论, 可见有必要用另一种方式来描述无机岩所含有有机质的丰度。

### 二、伽马分布的特点

经验表明, 用伽马分布来描述无机岩所含有有机质的丰度概率分布规律, 比较符合实际。伽马分布的密度函数如下:

$$y = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - \delta)^{\alpha-1} e^{-\beta(x-\delta)} \quad (1)$$

式中,  $x$  表示随机变量(例如有机碳丰度),  $\delta$  表示  $x$  的最小值( $\delta \geq 0$ ),  $y$  表示  $x$  的概率密度,  $\alpha$  和  $\beta$  是决定分布曲线形态的参数,  $\Gamma(\alpha)$  为伽马函数, 其数值视  $\alpha$  而定, 可查表得知[例如,  $\Gamma(1) = 1$ ,  $\Gamma(1.5) = \sqrt{\pi}/2$ ,  $\Gamma(2) = 1$ ,  $\Gamma(3) = 2$ ]。对伽马分布来说, 总体平均值  $a$ 、总体标准差  $\sigma$ 、峰值  $x_f$ 、三阶中心距  $\mu_3$  同  $\alpha$ 、 $\beta$ 、 $\delta$  之间, 存在如下关系:

$$a - \delta = \alpha / \beta \quad (2)$$

$$a - x_f = 1 / \beta \quad (3)$$

$$\sigma^2 = \alpha / \beta^2 \quad (4)$$

$$\mu_3 = 2\alpha / \beta^3 \quad (5)$$

$$x_f - \delta = (\alpha - 1) / \beta \quad (6)$$

$$(a - x_f) / (a - \delta) = 1/\alpha \quad (7)$$

由于对有机质丰度来说,  $a > \delta \geq 0$ ,  $x_f \geq \delta \geq 0$ , 所以: (1),  $\alpha$  不小于 1; (2) 平均值恒大于峰值, 三阶中心矩恒大于零, 分布曲线总是正偏的; (3), 当  $\alpha$  较大时,  $a - x_f$  比  $a - \delta$  小得多, 因而伽马分布曲线近似正态分布曲线。

如果我们采集了  $n$  块岩石样品, 并且获得了  $x_1, x_2, \dots, x_n$  总共  $n$  个实测数据, 我们就可以利用式(8)和式(9)算出样本平均值  $\bar{x}$  和样本标准差  $s$ :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (8)$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \left( \frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2 \quad (9)$$

统计理论指出, 样本平均值的标准差只有样本标准差  $s$  的  $n^{-0.5}$  倍, 样本方差  $s^2$  的数学期望等于总体方差  $\sigma^2$  的  $(n-1)/n$  倍。对有机质丰度数据来说, 样本容量  $n$  一般都比较小, 因而可以直接把样本的  $\bar{x}$  值和  $s$  值, 分别看作是总体的  $a$  值和  $\sigma$  值。于是, 我们可以按下列公式确定  $\alpha$ 、 $\beta$  值:

$$\alpha = (\bar{x} - \delta)^2 / s^2 \quad (10)$$

$$\beta = \sqrt{\alpha} / s \quad (11)$$

式中的  $\delta$  值需要根据样本的频率分布来估计。 $\delta$  值不可能小于零, 一般也不应该大于实测数据的最小值  $x_{min}$ 。为了便于进行有关的积分计算(例如求累积概率的计算), 最好这样来选择  $\delta$  值: 把它代入式(10)之后,  $\alpha$  值恰好是正整数。

至于伽马分布的分布函数  $F(x)$ , 其定义如下:

$$F(x) = P\{\delta \leq \xi < x\} = \int_{\delta}^x y dx$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_{\delta}^x (x-\delta)^{\alpha-1} e^{-\beta(x-\delta)} dx$$
(12)

具体说来，当 $\alpha = 1$ 时，

$$F(x) = 1 - \exp[-\beta(x-\delta)]$$
(13)

当 $\alpha = 2$ 时，

$$F(x) = 1 - [1 + \beta(x-\delta)] \exp[-\beta(x-\delta)]$$
(14)

当 $\alpha$ 为小数时， $F(x)$ 的数值需用数值积分方法算出。

已经知道密度函数 $f(x)$ 和分布函数 $F(x)$ 的具体形式，我们就能够把分布曲线（ $f$ 的图象）和累积分布曲线（ $F$ 的图象）画出来。

上述确定伽马分布各项参数的方法，其特点在于以样本平均值和样本标准差作为基本依据，故可称为“平均值—标准差法”。

### 三、有机碳丰度分布的一个实例

表1列出了湘中地区中泥盆统棋梓桥组底部至下石炭统梓门桥组底部255块海相灰岩样品（方解石含量均大于50%）的有机碳丰度实测数据。这255个实测数据组成的样本，其平均值为 $\bar{x} = 0.111\%$ ，标准差为 $s = 0.073\%$ 。图1是这套实测数据的频率分布直方图，它在一定程度上冲淡了样本的随机波动，并使人直观地看出有机碳丰度的峰值小于其平均值。

当我们利用上述样本特征来推断总体分布规律的时候，考虑到样本容量较大，可以认为， $a = \bar{x}$ ， $\sigma = s$ 。此外，由于最小实测值 $X_{min} = 0.01\%$ ，可知 $0 \leq \delta \leq 0.01\%$ 。根据式（10），若 $\delta$ 值取为0，则 $\alpha$ 值

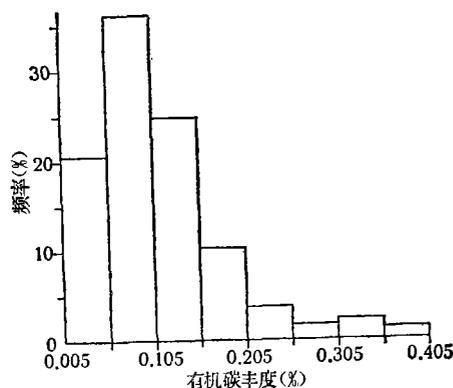


图1 湘中灰岩有机碳丰度的频率分布

应取2.3，若 $\delta$ 值取为0.01%，则 $\alpha$ 值应取1.9。为了便于进行有关的积分计算，现取 $\alpha = 2$ 。于是 $\delta = 0.0078\%$ 。至于 $\beta$ 的具体数值，则不仅与 $s$ 的数值有关，还与 $s$ 的单位有关。当 $\bar{x}$ 、 $s$ 、 $\delta$ 均以1%为单位时，根据式（11）， $\beta = 19.37$ 。这样一来，密度函数和分布函数分别是（图2、图3）：

$$y = 375(x - 0.0078) \exp[-19.37(x - 0.0078)]$$
(15)

$$F = 1 - [1 + 19.37(x - 0.0078)] \exp[-19.37(x - 0.0078)]$$
(16)

根据式（3）或式（6），可知峰值 $x_1 = 0.059\%$ 。将 $F = 0.5$ 代入式（16）中，可算出中位值 $x_z = 0.094\%$ 。

式（15）和图2所反映的不是255块样品的有机碳丰度分布情况，而是整个取样区间（棋梓桥组底部至梓门桥组底部）的全部灰岩的有机碳丰度分布规律。因此，如果给定了一个“灰岩生油岩有机碳标准”，我们就能够很容易地根据式（16）而把达标灰岩的分率（ $1 - F$ ）和不达标灰岩的分率（ $F$ ）算出来。例如，如果把标准定为0.15%，那么达标灰岩将占灰岩

湘中255块灰岩样品有机碳丰度表

表1

丰度 %	样品数	累计样品数	丰度 %	样品数	累计样品数
0.01	9	9	0.18	9	223
0.02	7	16	0.19	6	229
0.03	14	30	0.20	4	233
0.04	12	42	0.21	2	235
0.05	10	52	0.22	2	237
0.06	15	67	0.23	2	239
0.07	19	86	0.24	1	240
0.08	21	107	0.25	2	242
0.09	18	125	0.26	1	243
0.10	19	144	0.27	1	244
0.11	13	157	0.30	2	246
0.12	10	167	0.31	3	249
0.13	11	178	0.33	2	251
0.14	16	194	0.34	1	252
0.15	13	207	0.37	1	253
0.16	1	208	0.38	1	254
0.17	6	214	0.40	1	255

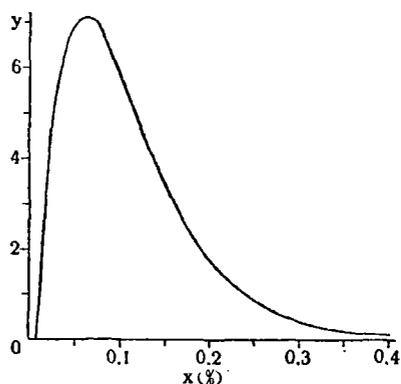


图2 湘中灰岩有机碳丰度分布曲线图

总量的23.9%，如果把标准定为0.20%，则达标灰岩仅占灰岩总量的11.4%。

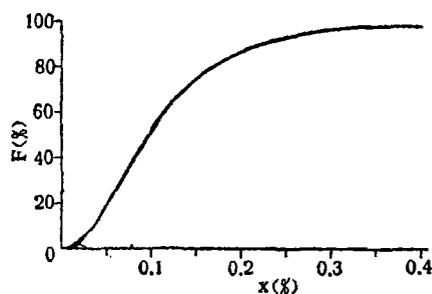


图3 湘中灰岩有机碳丰度累积分布曲线图

四、氯仿沥青丰度分布的一个实例  
采自湘中地区棋梓桥组底部至梓门桥

组底部的169块灰岩样品，其氯仿沥青丰度（表2），样本平均值 $\bar{x} = 23\text{ppm}$ ，样本标准差 $s = 20\text{ppm}$ ，最小实测值 $x_{\min} = 0$ 。

根据式（10）和（11）， $\alpha$ 应取值 1.32， $\beta$ 应取值0.0575。于是，密度函数为：

$$y = 0.0256x^{0.32}e^{-0.0575x} \quad (17)$$

湘中 169 块灰岩样品氯仿沥青丰度表

表2

丰度 (ppm)	样品数	累计样品数	丰度 (ppm)	样品数	累计样品数
0—9	53	53	50—59	13	163
10—19	37	90	60—69	2	165
20—29	24	114	70—79	1	166
30—39	20	134	80—89	0	166
40—49	16	150	90—99	3	169

峰值为 $x_f = 5.6\text{ppm}$ （表3，图4）。

至于分布函数，应为：

$$F = 0.0256 \int_0^x x^{0.32} e^{-0.0575x} dx \quad (18)$$

由于 $\alpha$ 是小数，F值需用数值积分方法算出（近似的计算结果已在表3列出）。但数值积分是一件十分麻烦的事情。为了简化计算过程，可取 $\alpha = 1$ ，于是根据式（10）和（11）， $\delta = 3$ ， $\beta = 0.05$ ，而分布函数

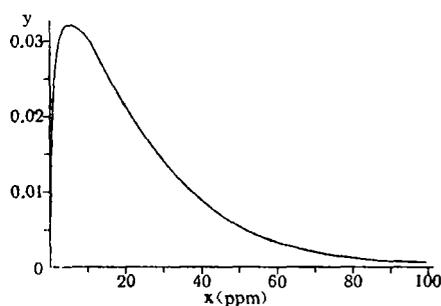


图4 湘中灰岩氯仿沥青丰度分布曲线图

湘中上古生界灰岩氯仿沥青丰度表

表 3

x (ppm)		5.6	10	20	30	40	50	60	70	80	90	100
样 本	累积样品数	36	53	90	114	134	150	163	165	166	166	169
	累积频率 (%)	21.3	3.14	53.3	67.5	7.93	88.8	96.4	97.4	98.2	98.2	100.0
总 体	$y \times 10^4$	322	301	211	135	84	51	30	18	10	6	4
	F (%)	15	29	55	72	83	89	93	96	97	98	99

注：由于组区间为左闭右开区间，故累积样品数累积至 $x-1$ 为止。

成为:

$$F = 1 - \exp[-0.05(x-3)] \quad (19)$$

不难检验,用式(19)算得的结果,同表3所列F值,相差不多。

### 五、问题讨论

如上所述,用伽马分布来描述无机岩石中有机质丰度的分布规律,有它的实用价值和方便性。如果人们有必要确定有机质丰度的概型(例如当应用蒙特卡罗法计算油气生成量的时候),可把它试为伽马型。但是,这种数据处理方法的提出,是以经验为基础而不是以逻辑推理为基础,我们难于回答“为什么会遵循伽马分布”这个问题。我们只能说,对于那些最小值 $\delta \geq 0$ 而且具有正偏分布图象的随机变量来说,用伽马分布来描述它们的分布规律,比较合适。

用任何一种分布律(包括伽马分布、正态分布等等)来描述随机变量,所依据的样本必须是随机样本,而且必须来自同一个统计总体,否则就会得出错误的结

论。举有机碳丰度为例,我们知道,灰岩的有机碳丰度偏低,其平均值、中位值和峰值一般都不大于1%,而油页岩的有机碳丰度则可能很高(例如澳大利亚二叠系油页岩可达81%),如果我们把灰岩有机碳丰度实测数据同油页岩的同类数据混合起来进行统计,我们将发现频率直方图是多峰的,甚至是奇形怪状的,因而这个混合样基本不符合伽马分布规律。进一步说,即使对于沉积环境、地质年代都相同的同一种岩石(例如灰岩)而言,如果取样方法不是随机取样(例如专选有机碳丰度高的灰岩作样品),那么所得到的样本的代表性也是很低的,它反映不了统计总体的分布特征。

(收稿日期:1983年1月7日)

### 参 考 文 献

- [1] 林少宫,基础概率与数理统计,人民教育出版社,1963。

## PROBABILITY DISTRIBUTION OF ORGANIC MATTER ABUNDANCE IN INORGANIC ROCKS

Yang Wenkuan

(Comprehensive Research Institute of Petroleum Geology,  
Ministry of Geology and Minerals)

### Abstract

It is shown by experience that the abundance of organic matter in inorganic rocks accords with the  $\Gamma$ -distribution, which is denoted as:

$$y = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - \delta)^{\alpha-1} e^{-\beta(x - \delta)}$$

where  $y$  is the probability density,  $x$  the organic matter abundance,  $\delta$  the minimum value of  $x$ , and  $\alpha$  and  $\beta$  the parameters determining the patterns of the distribution curve, respectively.

When the mean value of samples  $\bar{x}$  and the standard deviation  $S$  are obtained,  $\alpha$  and  $\beta$  can be defined by:

$$\alpha = (\bar{x} - \delta)^2 / S^2$$

$$\beta = \sqrt{\alpha} / S$$