第21卷 第1期
1985年1月

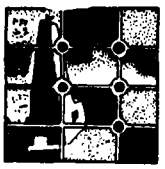
地质与勘探
GEOLOGY AND PROSPECTING

Vol. 21, No. 1
January, 1985

蒙特卡罗方法与矿产资源评价

杨永华

（长春地质学院）



蒙特卡罗方法也称统计试验法。它是根据统计抽样理论，通过对随机变量函数的概率模拟、统计试验来近似求解的方法。

矿产资源无论是作为地质过程的产物还是作为地质观测的结果，它都具有随机的性质。对资源量的计算，必然受概率法则的支配，因而是一定概率意义下的估计。由于蒙特卡罗方法能够正确地模拟随机变量的分布，再现它的取值规律，因而在矿产资源总量预测中被广泛用作资源估计的方法。

所谓资源量，是相对资源的位置而言，它是资源评价的重要内容。资源量包括质量和数量两方面特征，属于前者的有品位，属于后者的有矿床数、矿石量及金属量等，以下把它们统称之为参数，资源量就用这些参数来表达。

用蒙特卡罗方法计算资源量的过程由下列步骤组成：

（1）构造概率模型，建立资源量与参数之间的关系。例如金属量M与矿石量T和品位C的关系是

$$M = T \times C,$$

M，T和C在这里都是随机变量；

（2）建立这些参数的统计分布。一般这些参数的分布可以通过对样本实测值的统计和模拟来求得；

（3）产生随机数。随机数可在计算机上通过某种算法得到。

（4）抽样模拟，形成资源量的分布。每一个随机数，都对应参数的一个抽样，多次抽样的结果，得到一系列资源量$m_1 = c_1 \times t_1$，$m_2 = c_2 \times t_2$，……，统计这些m的取值规律，就得到资源量的分布。

（5）研究预测区成矿条件，用资源量分布模型估计预测区的资源，从而做出评价。

下面具体研究这些步骤的实现方法。

建立概率模型

蒙特卡罗方法的应用很广泛，只要能构造出适当的概率模型，几乎可以模拟任何问题。因此，构造概率模型是关键的一步。不同的模型，有不同的预测作用，要根据具体的工作目的加以研究。目前在矿产资源总量预测中，常用的几种模型是

（1）随机变量乘积模型；

（2）随机变量和模型；

（3）随机变量混合模型。

这里的随机变量都是指与资源量有关的参数，建立模型就是要在各个参数的分布已知的条件下，通过参数之间的不同关系，求得资源量的分布规律。不同的模型，反映不同性质的问题，因此有不同的适用范围。举例说，如果某人对某区域内一个矿带中产出的矿床数、矿石量及品位的统计或估计值分别看作三个随机变量N，T和C，那么，对于这个矿带来说，他对金属量M的研究就可以用N，T，C三者乘积来进行，这就是第（1）种模型。假如他所研究的区域内不止一个矿

带，譬如两个，而每个矿带中的矿床数、矿石量和品位有不同的分布，那么他对整个地区矿床数N或矿石量T的估计就可以通过和式$N = N_1 + N_2$或$T = T_1 + T_2$来实现，显然这属于第（2）种模型。最后，对于整个地区品位这一随机变量来说，则应当由两个矿带的品位C_1和C_2混合而成，这就是第（3）种模型。由这个例子可以看出，各种模型有不同的针对性，在研究一个较为复杂的问题时，可把它先分解为几个简单的阶段，建立适当的模型，然后根据各个阶段之间的关系，综合成一个统一的过程。这样，在一个问题中，概率模型可能不止一种，复杂的情形常是几个模型联合起来使用。

下面研究一个具体例子，它比较简单，仅涉及乘积概率模型，通过它来讨论一下构造概率模型的方式和参数的使用条件。

实例：研究某种与基性—超基性岩有关的矿产资源。在某个矿区内有29个已知矿床，它们的矿石量、品位数据列入表1。此外还知道矿区外围岩体的分布情况，现在把这些未知岩体作为研究对象，来估算预测区即矿区外围的资源潜力。

显然，这个问题可通过已知矿床建立资源量概率模型，然后把它推广到预测区去来解决。

用金属量来表示资源，建立概率模型$M = T \times C$，这里M是一个随机变量。对于模型区的29个矿床来说，M不再是已知储量。由于M是根据已知矿床的参数T和C随机相乘得到，而且又受T和C的分布制约，所以它代表的是模型区矿床的理论储量，它的规律性表现在随机变量M的分布上。当把这个模型推广到预测区去时，相当于假定了这样一个前提：预测区内“矿床”的资源量也服从M的分布。然而，问题是在预测区内取出一个岩体时，它的资源量能否用M的分布来直接表示？回答当然是不能。因为M表现的是样品为“矿床”条件下的资源量取值规律，而我们现在并未肯定这个待测岩体是否为“矿床”。但是，我们可以通过某种方法在预测区对各个岩体给出一个有利性指标，把预测对象调整到“矿床”的意义上来，使得模型区与预测区可直接类比，从而得到比较合理的评价结果。根据这种想法，把前面的模型改写成

$$M = T \times C \times L,$$

其中L是预测样品的类比系数，它是定义在（0，1）上的一个数，在我们的问题中，L表示预测岩体在多大程度上可能成矿。

矿石量及品位数据表　　表1

矿床	矿石量（万吨）	品位（%）	矿床	矿石量（万吨）	品位（%）
1	269123.49	2.16	16	7948.26	0.70
2	244248.50	0.57	17	15630.12	0.50
3	127419.46	0.64	18	4341.09	1.04
4	34581.98	0.83	19	49423.80	0.33
5	159807.79	0.30	20	51815.89	0.32
6	49744.56	0.43	21	1859.30	0.31
7	50307.24	0.28	22	67.96	0.32
8	11829.96	0.73	23	4349.25	0.25
9	14806.47	0.47	24	489.29	0.25
10	31714.19	0.30	25	209.31	0.26
11	16638.59	0.38	26	1429.82	0.38
12	10212.58	0.37	27	4349.25	0.25
13	13591.41	0.30	28	1209.24	0.50
14	8247.27	0.70	29	1630.9?	0.25
15	8872.47	0.40			

概率模型的构造与所研究问题的性质有密切关系。上面是以基性—超基性岩体为研究对象来建立预测模型的例子。下面考虑以单元（网格）为研究对象来建立预测模型的方法。由于单元中的矿床数N也是一个随机变量，它可能取0，1，2，3……等离散值，因此对于模型单元来说，其资源量M便与三个参数N，T，C有关；而对于预测单元来说，还需要考虑类比系数L，此时建立的预测模型就是

$$M = N \times T \times C \times L,$$

其中N，T，C的意义如前述，它们都来自模型区，而L是在预测单元中定义的一个随机变量。

上面讨论的两个模型都是矿床模型，按公式进行一次抽样，得到的是一个预测岩体或预测单元的随机储量。如果需要求出整个预测区的资源量，则可按下述方法考虑：设预测样品共有k个，将抽得的储量值每k个相加，就得到整个预测区的一个随机储量。这样重复下去，便有一系列资源量的取值，据此就可以做出预测区资源量

的分布。

构造概率模型的办法很多，根据问题的性质和资料水平的不同，可以有不同的模型。上面是通过对各个矿床的研究来求整个预测区的资源量，实际上还可以从其它途径入手，通过各个已知矿田之间关系的研究来求预测区资源量。

如某些类型矿床在空间分布上具有成群产出的特点。在一个矿田范围内，聚集着若干规模不等的矿床，其中最大的矿床，我们把它叫做主矿床。主矿床往往占有矿田总储量的很大比例。这样，只要知道了这种储量分配关系的规律性，那么对一个未知矿田的资源预测就可以通过主矿床与矿田的关系来研究。假设有一批同类矿田储量的资料，包括各矿田的总储量、矿田中主矿床的矿石量、品位及金属储量。通过这些资料，可以算得主矿床占矿田储量百分比，用字母E来表示，于是可以建立这样一个模型：

$$M = T \times C / E$$

其中T是主矿床的矿石量，C是品位，二者相乘是主矿床的金属量，再除以主矿床占矿田储量百分比E，就得到矿田的总储量M。在这个模型中，每次抽样，得到的是一个矿田的随机储量，多次抽样，就得到一系列矿田储量，根据这些值，最后就可以做出预测矿田的资源量分布曲线。

这个模型要求的条件较宽，对于工作程度不高，仅发现个别矿床甚至尚未发现矿床的远景区也可使用。

资源参数分布的模拟

除人们直接给出的资源参数分布以外，一般由实测数据建立统计分布，有两种方法：一是选用合适的已知分布律来拟合，二是用数学方法构造分布函数。下面分别研究这两种方法。

（1）首先对参数的原始数据进行整理，分组求频数，做出频率直方图，根据直方图的峰度、偏度等特征，选用已知的分布律来代表参数的分布。例如，据前人的研究，某些矿床点的分布服从负二项或普阿松分布，稀有金属矿床中元素含量和矿石量服从对数正态分布，等等。但是，这些结论都是经验性的，不一定具有普遍意义，在应用中只能作为参考。在实际工作中，我们可以根据直方图的形态用多种分布来试验，选择拟合程度较好的来使用。这里还应注意，直方图的形态不是固定不变的。分组的不同，样品的大小，都可能引起分布律的变化。用已知分布律来代表资源参数的分布，给我们研究问题带来很大方便。它的缺点是由于事先难以预料资源参数服从何种分布，因此在编制计算程序时需要考虑各种可能的情况，比较麻烦。

仍以前例中29个已知矿床的储量资料为例。将表1中矿石量数据取对数，用0.423为组距，分为9组，做出频数表（表2）并画出相应的频率直方图（图1）。

矿石量频数表　　表2

组中值										
	对数	2.044	2.467	2.890	3.313	3.737	4.160	4.583	5.006	5.430
	真数	110.74	293.29	776.78	2057.31	5461.35	14464.4	38308.9	101461.2	269153.5
频数		1	1	2	3	5	7	6	2	2

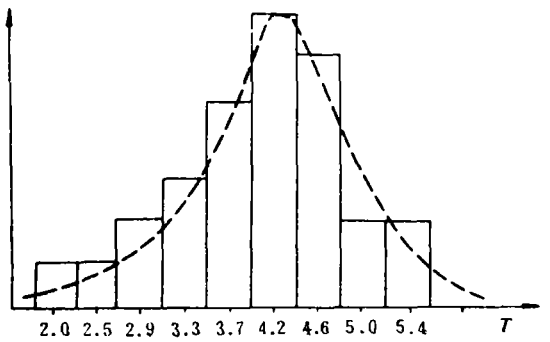


图1　矿石量频率直方图

根据直方图单峰、对称的特征，用正态分布来拟合它。为此求得（对数）算术平均值$\bar{x}$ = 4.014，标准差σ = 0.804。在表3中给出实际频率和参数（a = 4.014，σ = 0.804）的正态分布理论概率。从表3中可以看出理论概率和实测频率很接近，故认为用正态分布来拟合直方图是合适的。严格地讲，理论分布曲线的选配还需要经过检验，只有在理论分布与实测结果没有显著差

正态分布拟合矿石量概率表　　表 3

组中值(对数)	2.044	2.467	2.890	3.313	3.737	4.160	4.583	5.006	5.430
实测频率	0.034	0.034	0.069	0.103	0.172	0.241	0.207	0.069	0.069
正态理论概率	0.007	0.020	0.054	0.111	0.171	0.208	0.190	0.129	0.070

异时才能使用。

（2）构造概率分布函数 $F(x)$，也是在建立参数的频率直方图基础上进行。具体作法是：寻找合适的函数 $f(x)$，用它来拟合频率直方图，$f(x)$ 须满足作为密度函数的条件，即：$f(x) \geqslant 0$；$\int_{-\infty}^{\infty} f(x)dx = 1$。这是一个曲线拟合问题，数学上有很多方法实现。这里介绍一个较为常用的方法：用样条函数逼近直方图来构造 $f(x)$，有了 $f(x)$，自然就可以得到 $F(x)$。二次样条函数的计算公式为

$$f(x) = \sum_{j=0}^{N} y_j \Omega\left(\frac{x - x_j}{h}\right),$$

其中 y_j 为直方图每个柱的高度，h 为柱的宽度，$\Omega(t)$ 为基本样条函数：

$$\Omega(t) = \begin{cases} 0 & \text{当 } |t| \geqslant \frac{3}{2} \\ -t^2 + \frac{3}{4} & \text{当 } |t| < \frac{1}{2} \\ \frac{1}{2}t^2 - \frac{3}{2}|t| + \frac{9}{8} & \text{当 } \frac{1}{2} \leqslant |t| < \frac{3}{2} \end{cases}$$

下面以29个矿床的品位为例，说明 $F(x)$ 的求法。将表 1 中品位数据取对数，以0.24为组距分为 6 组，做出统计频数表（表 4）及直方图（图 2）。由直方图看出，原始数据取对数后图形仍不对称，故不能用对数正态分布来描述品位的分布状态。我们采用二次样条函数来拟合，利用公式可求出任意点的概率密度 $f(x)$，积分即得到任意区间的概率。为便于比较，将算得的各组理论概率列在表 5 中。

品位频数表　　表 4

组中值	对数	-0.517	-0.347	-0.176	-0.006	0.164	0.334
	真数	0.304	0.449	0.666	0.986	1.458	2.157
频数		13	8	5	2	0	1

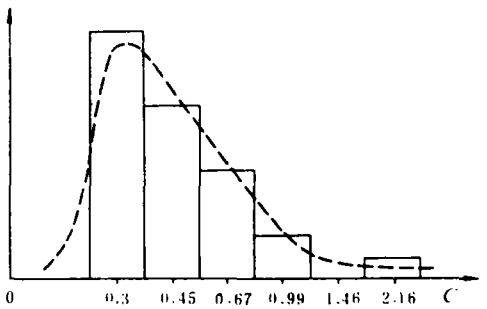


图 2　品位频率直方图

样条函数拟合品位概率表　　表 5

分组	1	2	3	4	5	6	计
实测频率	0.448	0.276	0.172	0.069	0	0.034	1
理论概率	0.427	0.285	0.172	0.073	0.013	0.030	1

构造函数 $f(x)$ 来拟合直方图的方法优点是算法统一，便于编制程序，而且应用中无需事先知道随机变量服从何种分布；缺点是 $f(x)$ 本身反映不出分布的性质，因而难以根据它对随机变量进行理论研究，但这并不影响蒙特卡罗方法的使用。

随机数与抽样

产生随机数的方法有多种，在蒙特卡罗方法抽样模拟中所使用的随机数是由计算机上用某种数学方法计算出来的。由于它受计算机字长限制而具有一定的周期，因此不是真正的随机数，这样产生的随机数被称作“伪随机数”。在应用中，可选择较好的算法使“伪随机数”有尽可能长的周期来保证抽样的随机性。在计算机上产生“伪

随机数”经常用的方法是乘同余法，计算公式为

$$x_{n+1} \equiv \lambda\ x_n\ (\mathrm{mod}M),$$

这是一个递推公式，给定x_0以后，就可以逐个计算x_1，x_2，…，譬如已计算了x_n，那么将x_n乘以常数λ，用M来除，取其余数即为x_{n+1}。例如取$M=59$，$\lambda=2$，$x_0=14$，算得随机数为：

14 28 56 53 47 35 11 22 44 29 58 57
55 51 43 27 54 49 39 19 38 17 34 9
18 36 13 26 52 45 31 3 6 12 24 48
37 15 30 1 2 4 8 16 32 5 10 20
40 21 42 25 50 41 23 46 33 7

这是一个整随机数列，它在（1，58）上均匀分布，要想把它变换到（0，1）上去，只要将每个数减去1再除以57即可。在抽样过程中使用的就是（0，1）上均匀分布的随机数，当然这里面不限于58个数了。

有了随机数后，就可进行抽样了。所谓抽样，就是在已知的某个随机变量分布情况下，通过取随机数实现在该变量中一次次取值的过程。例如准备在品位C的分布下抽样，为此将C的取值区间分成n份，则品位C落在各个小区间上为事件C_1，C_2，…，C_n，而相应的概率为p_1，p_2，…，p_n。若令$p^K=\sum_{i=1}^{K} p_i$和$p^0=0$，则p^K表示了前K个概率之和，显然有$p^K-p^{K-1}=p_K$。现取（0，1）上均匀分布的随机数r，若r落在（p^{K-1}，p^K）上，则说事件C_K发生（即取得了一个品位为$C=c_K$的样品），其概率为p_K。这样，就使一个随机数r与C的一个取值c_K对应起来，从而完成了一次抽样。

不同的抽样方式，产生不同的随机结果。抽样方法的使用取决于两个条件：一是概率模型的形式，二是参数的性质。仍以岩体的预测模型$M=T\times C\times L$为例，式中参数T，C的分布已经在前面给出，又假设类比系数L的分布也已做出，那么对每个预测岩体金属储量估计值可由下述抽样过程模拟出来：先取一个（0，1）上均匀分布的随机数r_1，在T的分布中抽得一个矿石量t_1，然后再取随机数r_2，在C中抽得一个品位值c_1，第三次取随机数r_3，在L中抽得l_1，将这3个值相乘，便得到一个预测岩体金属量的随机值$m_1=t_1\times c_1\times l_1$，于是完成了一轮抽样。依此做下去，譬如进行了1000轮抽样，便有1000个金属量的取值m_1，m_2，…，m_{1000}，把这1000个数分组求频率，整理得M的分布。这是对单个预测岩体所做的资源估计，使用的是乘积概率模型。随机变量和的模型及混合模型抽样方式与此类似，只是每轮抽得的参数值之间不再作乘积。举一混合模型的例子。有A，B两位专家对某种与基性—超基性岩有关矿床储量给出了分布M_1和M_2，现在要对一见矿岩体的资源量M作出估计。假设专家的水平不同，A的信任系数是0.6，B是0.4，于是可以设计这样一个抽样程序：连续取随机数r_1，r_2，r_3在M_1中抽3次得m_1，m_2，m_3，又取随机数r_4，r_5在M_2中抽得m_4，m_5，将这5个金属量抽样值依次排列，算作一轮抽样的结果，按此方式进行1000轮，便有5000个随机储量，经整理最后即可得到M的分布。

参数的性质，对抽样方法的使用有很大影响。前面的讨论，都是在资源参数之间相互独立的假设条件下进行的。但是，这个条件在某些情况下不能满足。例如，在研究某些矿床时发现，主矿床储量与它在矿田中所占比例存在着一定的关系。表6给出12个国内较大的某种矿床的储量资料。

12个矿床储量关系表　　表6

矿床	储量（万吨）	占矿田百分比	矿床	储量（万吨）	占矿田百分比
1	4399.00	99.9	7	40.9	73.3
2	177.45	78.5	8	24.9	53.3
3	161.68	73.6	9	24.4	59.8
4	124.99	100.0	10	17.4	57.6
5	107.9	93.2	11	21.8	87.9
6	41.5	74.0	12	12.7	68.2

由表6看出，主矿床的储量越大，其所占矿田总储量的百分比也越高。其中主矿床储量在100万吨以上的，所占百分比为70～100，主矿床在100万吨以下百分比为50～90。这说明主矿床

储量和百分比这两个参数之间不完全独立。在这种情况下，就要分别建立百分比E的两个分布：E_1对应$M \geqslant 100$，E_2对应$M < 100$。如果模拟矿田储量使用概率模型$M = T \times C / E$时，当取随机数r_1得t_1、取r_2得c_1后，乘出m_1'，视m_1'取值若大于100，则取r_3在E_1中抽取，否则在E_2中抽取，最后将m_1'除以抽得的e_1即可得到一轮抽样结果m_1。

根据国外一些学者的研究，某些类型矿床的矿石与品位呈反消长关系，小型矿床往往较富而大型矿床则较贫，例如斑岩铜矿就是这样。这个规律使得矿石量和品位两个资源参数之间不独立，因此在模拟中对抽样方法要作类似前面的处理，这样才能保证结果的可靠性。

最后要强调一点，对于一个概率模型中的各个参数的一轮抽样，取随机数的过程不能化简。就是说，有n个参数，就必须取n个随机数，或者说一轮抽样应由n次抽样构成。这是因为在每个分布中，一个随机数对应一个确定值，例如$M = T \cdot C$中，r_0和t_0，c_0都是确定的对应关系，如果只用一个随机数同时在T和C中抽样，那就破坏了T和C之间的独立性，而使模拟结果与实际情况不符。

资源量的估计

一系列随机抽样的结果，得到一系列资源量的取值，对这些值进行统计便可得到资源量的概率分布$p(x)$和概率分布函数$F(x)$，x代表资源量，它或者是金属量，或者是矿石量、品位等。一般资源量的表达都使用$F(x)$，它实际上就是累积概率，为便于解释，又将它作一变换：$\overline{F}(x) = 1 - F(x) = 1 - p(\xi < x) = p(\xi \geqslant x)$，这是定义在$(0, \infty)$上的单调减函数，其图形如图3（左）。图中，若横轴$x$代表金属量(吨)，纵轴代表累积概率，则查得纵轴上0.75对应曲线之下的x_0，读作：金属储量大于x_0吨的可能性不超过75%。

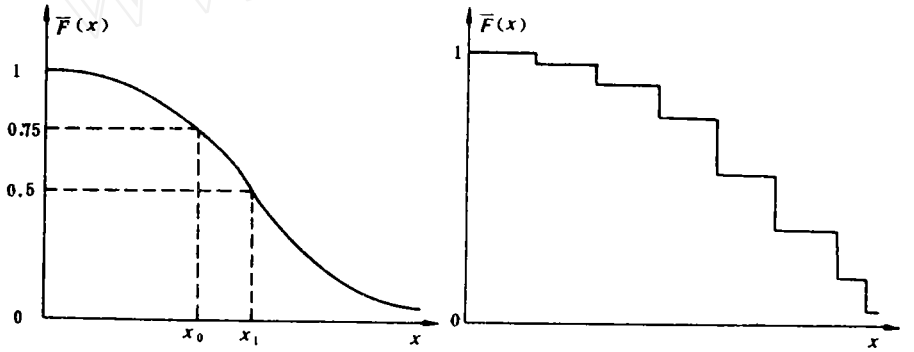


图3　资源量累积概率$\overline{F}(x)$分布曲线

对于离散取值的资源参数如矿床数，其分布形式是一条阶梯状的折线，类似对连续曲线的分析，可根据它读出一定概率下的不小于某数的矿床数估计值（图3右）。

预测模型的应用

使用蒙特卡罗方法对资源量的预测过程，是应用随机抽样的理论对资源量的分布进行模拟的方法来完成的。它所依据的原始资料，主要是矿床数、矿石量、品位及储量等数据。这些数据可以来源于实际观测，也可以由人根据经验估计给出。作为一种计算技术，对于前者，它属于矿床模型法；而对于后者，则属于主观概率法。不论哪种方法，都须注意一个资源含义的问题。由于以往的工作基本上是研究“储量”，而对资源预测来说，这些储量数据是不够的。因为依靠储量资料所作出的估计，并非全部资源，它没有包括“潜在”的那一部分在内。通常这类资料不易直接得到，它们很多未经整理，分散在各种原始资料中。由于过去对这一部分资源的特点和规律注意得不够，因此即使是专家的估计也可能出现偏差，所以应当有目的地搜集和整理这些资料，在计算中加以利用。除掉增补纯粹属于“资源”的那类矿点、矿化点的资料以外，对原有的矿床数据也需进行改造。这一般可通过调整参数的临界值来实现。

参 考 文 献　（从略）