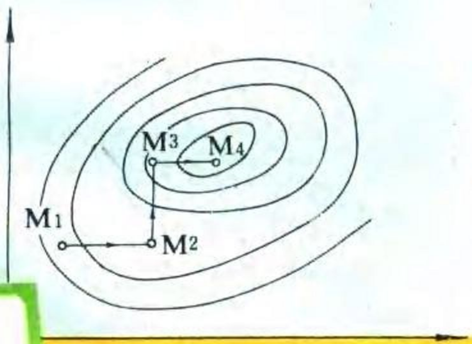
勘探钻进试验设计

〔苏〕Д.Н.巴什卡托夫 著

鄢泰宁 顾家林 译

汤凤林 校



中国地质大学出版社

内 容 提 要

书中阐述了勘探钻进过程中提供工艺信息的原理。研究了在多因素作用的过程中，如何设计、实施对比试验，并对其结果进行评估的问题；还研究了如何使用多元分析方法来使钻进规程优化的问题。书中列举了一些勘探钻进中的实例。

本书的读者对象是从事研究、试验钻探技术和改进钻进工艺的工程技术人员。也可供高等院校“探矿工程”专业的学生使用。

Д.Н.БАШКАТОВ
ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА
В РАЗВЕДОЧНОМ БУРЕНИИ

МОСКВА “НЕДРА” 1985

勘探钻进试验设计

〔苏〕Д.Н.巴什卡托夫 著
鄢泰宁 顾家林 译
汤凤林 校
责任编辑 刘先洲
责任校对 杨 霖

*

中国地质大学出版社出版
湖北省新华书店发行
湖南省地矿局402队彩色印刷厂

*

开本 787×1092 1/32 印张 7.875 字数176千字
1989年5月第1版 1989年5月第1次印刷
印数1—2000册 定价：1.65元
ISBN 7-5625-0297-8/TD·5

译　序

钻进过程是一个受众多人为因素和随机因素影响的复杂系统。通过数据处理可以找出钻探效率、质量和成本与这些因素之间的规律。但是如果仅仅被动地处理大量实测数据，而不对试验进行设计，不仅会盲目增加试验次数，浪费时间和资金，而且还不一定能得到满意的结果。近年来，随着技术进步和计算机的普及，探矿工作者越来越希望在寻找最优工艺参数和配方，建立生产过程数学模型等问题中，能减少试验次数（在目前钻机现场承包的情况下更不允许安排大量试验），主动地把试验的安排，数据的处理，乃至数学模型的建立统一加以考虑。

近十年来，苏联学者在减少试验次数，进行试验设计和数据处理，建立模型方面做了许多工作。Д．Н．巴什卡托夫（Башкатов）教授著的《勘探钻进试验设计》把这方面的成果和方法系统化了。我们把这本书译出，供国内同行在由凭经验打钻走向科学打钻的进程中借鉴、参考之用。

书中涉及了不少概率统计和矩阵方面的知识，同时列举了许多生产和科研的实例。如果读者是实际工作者，对有关的工程数学知识比较陌生的话，建议把书中实例的数据处理过程与算法对照起来阅读，将收到事半功倍的效果。

全书由顾家林同志译第一、二章，鄢泰宁同志译第三、四、五章并负责整理全书译稿。在翻译过程中，我们对原书中的错漏之处作了修正，凡属印刷错误均未加注明。

原书是汤凤林教授1986年从苏联带回来的，译稿完成后又承蒙他认真审校。在此，向他表示衷心感谢。

由于译者水平有限，译稿中难免有不少缺点错误，敬请读者批评指正。

译　者

1987年11月

前　言

在确定钻进中各种指标和评价所用的技术手段及工艺过程中，试验的作用和意义是很大的。通过试验可以检验新技术手段和工艺过程的有效性和可靠性，可以确定新技术的生产效率和技术经济指标等等。

通过试验应该能回答试验后结果如何的问题。但是，由于很多因素将对试验结果产生影响，因此，试验一般要重复多次。试验结果的可靠性，以及试验延续时间的长短、费用的多少，都与是否善于正确设计和安排试验有关。

试验结果是用概率论和数理统计的方法评价的。这些方法不仅能判定试验结果，而且能设计试验的次数和精度，还可根据在进行试验阶段得到的试验结果修正这些数值。

直到二十世纪初，在经典学科中仅仅讨论了组织得很好的系统，并在概率论和数理统计的基础上发展了单因子试验方法。这类问题带有抽样和误差的性质，一般归结为在所有其余因子不变的情况下，研究者变动某一因子并建立这样或那样的关系式。

在本世纪20年代初，由于英国数学家P．A．菲舍尔（Фишер）的工作，多元统计方法得到了广泛的应用。该方法规定在进行试验时，可按专门的计划同时改变诸因子的数值。

苏联和国外的研究者B．B．纳利莫夫（Налимов），

Ю．П．阿德列尔（Адлер），科赫连（Кохрен），科克斯（Кокс），芬尼（Финни）等人对发展这些方法作出了重要贡献。

1951年，鲍克斯（Бокс）提出了在多因子试验中寻找近似最优区域的方法，随后又在大量有关试验理论的著作中发展了这些方法。复杂的系统要通过逻辑分析、用控制论的方法研究，并用电子计算机来控制。

在钻探中，统计评价和试验设计方法的发展起步于50年代中期，它与В．С．菲多罗夫（Федоров）和В．Г．别利科夫（Беликов）的工作是分不开的。Е．А．科兹洛夫斯基（Козловский），А．Х．米尔扎德让扎杰(Мирзаджанзаде）及其学派的学生为在钻探中发展科学试验方法作出了最大的贡献。

然而，就所研究的问题只出版了几本专著，我们从中选用了А．Х．米尔扎德让扎杰[34, 35, 36]，Е．А．科兹洛夫斯基[18]，В．И．伊格纳托夫（Игнатов）[13]和В．Г．别利科夫[4]的著作。

还应该指出Г．Ц．图马尔金（Тумаркин），В．П．罗日科夫（Рожков），В．М．尔瓦切夫（Рвачев），С．С．科切尔茹克（Кочержук），А．П．托恩（Тонь），Ю．Г．索诺沃夫（Соловов），Ю．В．瓦杰茨基（Вадецкий），Р．М．埃格列斯（Эйгелес）等也进行了这方面的研究。

在70年代初，Е．А．科兹洛夫斯基在钻进过程的研究和控制方面开拓了新的研究方向——钻进过程的最优化。在А．А．波加尔斯基（Погарский），Р．Х．加菲亚杜尔林（Гафиятуллин），Н．И．捷列霍夫（Герехов），В．Д．布

特金（Буткин），В.А.齐甘科夫（Цыгаков），А.Е.特罗普（Троп），И.П.别特罗夫（Петров）等人的著作中，成功地发展了这一研究方向。试验是在实验室和生产条件下进行的；在许多情况下，为了进行研究，建立了能模拟某些条件和过程的专用试验台。试验可以是被动的，也可以是主动的。

被动试验就是研究者在进行观察和评价各种关系或参数时，对工艺过程不加干预。这类试验的实例有：确定指定矿区钻孔弯曲的规律性；评价机台和钻具的生产效率；在指定的条件下评价各种技术装备发生故障的次数等。

主动试验考虑工艺过程的变化，以便更充分地查明我们感兴趣的现象或事物的特性。比如在给定的矿区地质条件下所作的钻头试验并确定合理的钻进规程就是主动试验的实例。在这种情况下，研究者以一定的方式改变规程参数并确定如下的一些指标，象机械钻速、钻进效率以及每米进尺成本等等。

所谓极值试验在实践中得到了广泛的应用。进行这种试验时，为了探求极值，研究者变化几个可变因素。极值的寻求是连续进行的。在每次试验之后或者完成一批试验之后，都评价试验结果并确定试验的对策。

钻进过程是一个复杂的系统，其中包含了大量的彼此相关的随机因素。因此，其研究结果带有概率的特征。可以这么说，在充分可靠的理论和试验前提下建立和研究钻进过程的随机模型，即使不是唯一的，也是最有效的一种研究钻进过程的方法。

作者编写本书的目的是把钻探试验设计的各种资料归纳综合起来并使之系统化。主要着重于多因子对比试验的设计、实施以及试验结果的评价。所研究的问题都用勘探钻进的实例加以说明，这会帮助读者易于掌握本书的资料并将它用于实践活动

中去。

本书综合了钻探试验设计方面各种不同的资料，因此，作者将十分感谢所有对本书提出意见和希望的人。

目　　录

第一章　概率论的基本概念

让我们简单谈一下在以后讨论所提出的问题时要用到的定义和公式。

§1　总体与抽样

为了研究总体容量很大的随机变量的某种特性和最重要的性质，可以考察有足够容量的某个抽样，以便用此抽样来查明所研究的随机变量总体的特点。从中进行抽样的容量很大的总体称之为全部总体*。抽样有大抽样和小抽样之分。抽样的容量越大，其特征（估计），例如数学期望、方差等，就越接近于总体的特征。但是，在样本的容量足够大时，这种差别就与抽样容量的大小没有多大关系了。因此，在实际中，没有必要研究容量太大的样本，因为这样作会增加研究费用和研究时间。

例如，装备地质勘探企业的所有带液压给进的钻机，在某些假定条件下，可以看成是一个总体。为了分析它们的工作情况，没有必要考察全部总体，只要对构成一个样本的一组钻机的工作情况加以分析就足够了。样本的相对指标用样本数与总体容量的比值来表示。

样本的平均值与总体的平均值之间的差别，是有代表性的误差。这个误差随着样本容量的增加而减少。在统计学中，包

* 以下将总体和全部总体统称为总体——译注

含100—150个个体的样本称为大样本，包含20—30个个体的样本称为小样本。对每种形式的样本采用相应的处理方法。收集的试验结果资料构成了统计总体，该总体由数个具体定量指标 x_1，x_2，x_3，……，x_n组成。统计系列资料也可以用直方图表示。

研究者一般只掌握有限的资料信息。因此，抽样参数分布的统计评价问题是特别重要的。选择分布曲线，一般要考虑边界条件。理论曲线要与经验分布曲线对比，并要确定两者的相关程度。

按抽样资料估计的总体分布参数，将是有偏差的；因此，考虑这种偏差并随样本的增加而修正这种偏差是很重要的。

§2 指标的平均值及其数学期望

变量 x 的平均值用下式表示

$$\bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{n} \tag{1}$$

式中，$\bar{x}$——变量的平均值；x_i——随机变量值；n——试验的次数。

数学期望$M(x) = \mu$——是当测量的次数无限增加时，参数平均值趋近的一个极限。

离散型随机变量的数学期望，用随机变量所有的可能值与其对应概率值乘积之和的形式表示，即

$$M(x) = \sum_{i=1}^{\infty} x_i P_i, \tag{2}$$

式中，P_i——概率。

连续变量的数学期望为

$$M(x)=\int_{-\infty}^{+\infty} xf(x)\,dx \tag{3}$$

连续随机变量的最频值M_o表示了该随机变量的概率密度最大的数值，而对离散型随机变量来说，M_o是它的最大概率的值。

中位数M_e是概率为0.5的随机变量值。这时，在样本中遇到比M_e大的数和比M_e小的数的机会是相等的。对于对称型分布来说，M_o和M_e重合。

往往将原始资料划分为区段和组，于是平均值用下式计算

$$\bar{x}=\frac{\sum_{i=1}^{i=n} x_{mi}m_i}{n} \tag{4}$$

式中，x_{mi}——与给定组、给定区段相应的特征值；m_i——在给定组中的变量绝对频数（权）或者特征值x_{mi}的个数。

按（4）式还可以确定加权平均值。

§3 离 散 度

方差和均方差表征了随机变量对平均值的偏离程度。

随机变量的方差表征了变量围绕其平均值的离散程度

$$D=\frac{\sum_{i=1}^{i=n}(\bar{x}-x_i)^2}{n} \tag{5}$$

$\sigma=\sqrt{D}$值称为均方差（标准差）

$$\sigma=\pm\sqrt{\frac{\sum_{i=1}^{i=n}(\bar{x}-x_i)^2}{n}} \tag{6}$$

对于容量不大的试验（$n<30$—50）来说，要引进方差修正值$n/(n-1)$，这可得出所谓无偏估计

$$D=\frac{\sum_{i=1}^{i=n}(\bar{x}-x_i)^2}{n-1} \tag{7}$$

对于标准差来说

$$S=\pm\sqrt{\frac{\sum_{i=1}^{i=n}(\bar{x}-x_i)^2}{n-1}} \tag{8}$$

当n增大（$n>30$）时，$S\to\sigma$。

在近似计算中，对于$n<10$的样本来说，最大的S值可用变化幅度（$R=x_{max}-x_{min}$）来近似计算

$$S\leqslant\frac{R}{2}\sqrt{\frac{n}{n-1}} \tag{9}$$

算术平均值的均方差按下式计算

$$S_{\bar{x}}=S/\sqrt{n} \tag{10}$$

式中，S——样本均方差σ的无偏估计。

在计算方差估计$S^2\approx D$中，现有观测值随着试验次数增多

而分组，因此按下列公式进行计算是合理的（这些公式是根据公式（7）推导出来的），按照平均值和现有值的分组分别为：

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

$$S^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i{}^2-x\sum_{i=1}^{n}x_i\right);$$

$$S^2=\frac{1}{n-1}\left[\sum_{i=1}^{n}x_i{}^2-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2\right]$$

样本方差也可以按分组资料借助于频数表来计算

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(u_m-\bar{x})^2h_m \tag{11}$$

式中，u_m——第m个区间的中值；h_m——落在第m个区间中的频数；$\bar{x}$——按频数表计算得出的算术平均值，

$\bar{x}=\frac{1}{n}\sum_{m}^{n}u_m\ h_m$；$n$——总的测量次数。

随机变量的方差称为二次中心矩$M_2=\sigma^2$。

变异系数能评价不同的变量：

$$V=\sigma/\bar{x} \tag{12}$$

或者用百分数形式表示为

$$V=\frac{\sigma}{\bar{x}}100\%,\ \bar{x}\neq0$$

变异系数的均方值为

$$\sigma_V = \frac{V}{n}\sqrt{\frac{1}{2} + \left(\frac{V}{100}\right)^2} \tag{13}$$

在试验室条件下进行的钻具研究具有高度的准确性，保证$V<8\%$。在生产条件下的试验研究一般得$V=8—15\%$。如果各种因素对研究过程有很大的影响，则$V>15\%$。例如，如果钻进是在硬度和研磨性不均匀的岩石中进行的，则钻进指标严重摆动，于是$V>15\%$。当$V>35\%$时，研究的结论通常带有定性的性质。

不对称指标是

$$A = \frac{\sum_{i=1}^{i=n}(x_i-\bar{x})^3}{n\sigma^3} \tag{14}$$

随机变量对其数学期望分布的不对称程度，用三次中心矩来表示

$$M_3 = \sum_{i=1}^{i=n}(x_i - M(x))^3 P_i \tag{15}$$

或者
$$M_3 = \int_{-\infty}^{+\infty}(x_i - M(x))^3 f(x)dx \tag{16}$$

对于对称分布来说，$M_3=0$。

峰度指标是

$$E = \frac{\sum_{i=1}^{i=n}(x_i-\bar{x})^4}{n\sigma^4} - 3 \tag{17}$$

统计估计分为点估计和区间估计。在第一种情况下，是用某个数来估计未知特征或者未知参数的；而在第二种情况下，是用某个区间来估计的。所作的估计要符合无偏性和可靠性的要求。由于估计也是随机量，那么，理想的情况是它趋近于被估计的参数。如果在试验（观测）的次数n为任何数时，估计的数学期望等于被估计的参数（即$M\theta_n^*=\theta$,这里M表示数学期望；θ^*和θ分别为被估计参数的估计值和未知真值），那么，这种估计称为无偏估计。

如果对于无论多么小的正数ϵ（$\epsilon>0$）来说，当试验次数n无限增加时（$n\to\infty$），所作的估计相对被估计参数真值的绝对偏差小于数ϵ的概率都趋近于1，那么，这种估计称为可靠估计。可靠性的数学条件写成下列形式

$$P\ (\mid \theta_n^*-\theta\mid<\epsilon)\to 1$$

这里，P——概率的符号。

同一个参数可能有数个估计都满足无偏性和可靠性的条件。此时自然选取方差为最小的估计。这些估计称为有效估计。为了估计所观测的特征值，利用调和平均估计

$$\bar{x}_\Gamma=\frac{n}{\sum\limits_{i=1}^{i=n}\frac{1}{x_i}} \tag{18}$$

几何平均估计

$$\bar{x}=\sqrt[n]{x_1\cdot x_2\cdot\cdots\cdot x_n} \tag{19}$$

在数据分布很不对称时，算术平均估计将失去其有效性。在这种情况下，可以利用最大似然估计。对于对数正态分布来说，最大似然估计按下式计算