

提高估计的可信性

胡桂铭

(地矿部石油地质研究所)

不确定性问题,是地质科学量化一个带根本性问题,作者从概率的概念出发,提出了估算某个变量可能取值,以概率分布取代单值估算的观点,进一步对概率分布估算参数的可信性进行了讨论。

用概率分布来估计一个参数的可信性的方法很多,反馈法把数学计算的精确意义与人的思维判断有机地结合起来,应用反馈机制,本程序以夏普PC-1500袖珍计算机作为工作机。

在解决不确定性问题时,常常要估计某个变量的可能取值。而在这类问题中,往往是我们对所估计的变量知之不多。因此,这时估计一个分布,就比用一个单值估计要好得多。有人做过大量试验,结果表明:人们对自己估计的准确性往往评价太高,或者说,在给定一定的概率区间范围内,人们给出的估计值往往偏窄。而且,这种偏差,与一个人的专业知识并无显著关系,它是由于人们头脑中缺乏概率的概念引起的。

提高用概率分布来估计一个参数的可信性的方法很多,其中一个反馈法,即把所作的估计的含义充分引伸,看所引伸出来的结论与我们估计的原意是否一致,如不一致,就要修改原来的估计。例如,估计“一个勘探层的油藏油量有80%可能在 40×10^6 — $100 \times 10^6 \text{m}^3$ 的范围内”是什么意思呢?估计的人不一定意识到它的全部意义。如果我们假设:油藏油量分布是成正态分布的,取值的不确定性是对称的,也就是说,油藏有10%的可能大于 $100 \times 10^6 \text{m}^3$,而小于 $40 \times 10^6 \text{m}^3$ 的可能性也是10%。

在这种假设的前提下,“油藏油量为 40×10^6 — $100 \times 10^6 \text{m}^3$ 的概率为80%”这句话等价于一个均值为 $70 \times 10^6 \text{m}^3$,标准差为 23.4×10^6 的正态分布。这个勘探层中的油藏大于 $15.54 \times 10^6 \text{m}^3$ 的概率为0.99,而大于 $118 \times 10^6 \text{m}^3$ 的概率为0.02。一般来说,这样的估计范围是太窄了。

如果把上述的假设改为油藏油量成对数正态分布,则“油藏油量为 40×10^6 — $100 \times 10^6 \text{m}^3$ 的概率为80%”这句话表示一个对数均值为17.96,对数标准差为0.3575的正态分布,油藏大于 $27.5 \times 10^6 \text{m}^3$ 的概率为0.99,而大于 $131.79 \times 10^6 \text{m}^3$ 的概率为0.02。

本程序是用夏普PC-1500袖珍计算机进行反馈工作的。

一、功能

本程序适用于正态分布或对数正态分布。它可以

1.由给出的均值、标准差计算分布曲线。

2. 由给出的累计分布（大于）曲线上的两点，计算均值、标准差，并计算分布曲线。

二、计算过程简述

标准正态分布的分布函数为：

$$\Phi(u) = P(x \leq u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{x^2}{2}} dx \quad (u \geq 0)$$

此函数可用多项式逼近：

$$P(x \leq u) \cong \text{ERF}(u)$$

$$= 0.5 + \text{SIGN}(u) \times (0.5 - 0.39894228 e^{-\frac{u^2}{2}} \\ \times (0.319382Q - 0.356564Q^2 + 1.781478Q^3 \\ - 1.821256Q^4 + 1.330274Q^5))$$

式中SIGN(u)为符号函数

$$Q = \frac{1}{1 + 0.2316419|u|}$$

给出标准正态分布中的一个u值，可求出 $\phi(u)$ ；相反，给出一个 $\phi(u)$ ，也可求出u。

对于任一正态分布的任一取值x，可以表示为 $x = u \cdot \sigma + \mu$ 。显然，对于标准正态分布， $x = u$ 。本程序是由所给出的 $P(X \geq x)$ ，反求u。通过两个点 $(x_1, P(X \geq x_1))$ 、 $(x_2, P(X \geq x_2))$ 求得 u_1 、 u_2 ，解联立方程

$$\begin{cases} u_1 \sigma + \mu = x_1 \\ u_2 \sigma + \mu = x_2 \end{cases}$$

求得 μ 与 σ 。

由对数正态分布的两点求 μ 与 σ ，只需将x改为 $z = \ln x$ 即可，最后计算分布曲线时，取 $x = e^z$ 。

三、程序的使用

1. 使用本程序，需输入三个（组）数据，依次为（程序运行中均为输入提示）：
 - a. 此分布是什么型（正态——0；对数正态——1）。
 - b. 用什么原始资料求分布（ μ, σ ——0；分布曲线上的两点——1）。
 - c. 原始数据：按程序询问的项目输入 μ, σ 或 X_1, PX_1, X_2, PX_2 。
2. 执行本程序最后输出一个由27点组成的分布曲线，每点由两个数组成：累计频率及其对应的X值。

四、PC-1500机BASIC语言程序

```
3000 REM PROGRAM DIST FOR COMPUTING A NORMAL OR
LOGNORMAL DISTRIBUTION
3010 REM FROM MU AND SIGMA OR TWO VALUES ON A
CUMULATIVE DISTRIBUTION
3020 REM XI(X2); VALUE OF VARIABLE
3030 REM FXI, U; SEE p. 5—88, A HANDBOOK OF MATHEM-
ATICS IN BRIEF
3040 REM EDUCATION PUBLISHING OF SHANGHAI
3050 REM ERF; FUNCTION FOR CALCULATING F(u)
3060 DIM H(5), PS(27), S(27)
3070 DATA 0.319382, -0.356564, 1.781478, -1.821256, 1.330274
3071 DATA .99, .98, .96, .94, .92, .9, .85, .8, .75, .7, .65, .6,
.55, .5
3072 DATA .45, .4, .35, .3, .25, .2, .15, .1, .08, .06, .04, .02,
.005
3073 DATA -2.326345, -2.053747, -1.750686, -1.554773, -1.405072,
-1.281552, -1.036434
3074 DATA -.841621, -.674489, -0.52440, -0.38532, -.253347,
-.125662, 0
3075 DATA 0.125662, .253347, .38532, .52440, .674489, .841621,
1.036434
3076 DATA 1.281552, 1.405072, 1.554773, 1.750686, 2.053747, 2.57583
3080 FOR I=1 TO 5
3090 READ H(I)
3100 NEXT I
3101 FOR I=1 TO 27
3102 READ PS(I)
3103 NEXT I
3104 FOR I=1 TO 27
3105 READ S(I)
3106 NEXT I
3107 INPUT "NORMAL OR LOGNORMAL? (0/1)", AA
3108 PAUSE "WHAT ARE BEING INPUTTED?"
3109 INPUT "MU, SIGMA OR X, P(x)? (0/1)", BB
3110 IF BB=0 GOTO 3373
```

```
3112 INPUT "X1=? ", X1
3115 INPUT "P(X1)=? ", P1
3120 INPUT "X2=? ", X2
3125 INPUT "P(X2)=? ", P2
3130 IF AA=0 GOTO 3140
3132 X1=LN(X1)
3134 X2=LN(X2)
3140 F1=1-P1
3150 F2=1-P2
3235 PAUSE X1, " "; P1
3236 PAUSE X2, " "; P2
3240 FX=F1
3245 GOSUB 3400
3260 U1=U
3270 FX=F2
3280 GOSUB 3400
3290 U2=U
3300 DELTA=U2-U1
3310 IF DELTA<>0 GOTO 3350
3315 WAIT
3320 PRINT "SORRY, DELTA="; DELTA
3330 GOTO 3391
3350 MU=(X1*U2-X2*U1)/DELTA
3360 SIGMA=(X2-X1)/DELTA
3365 WAIT
3370 PRINT "MU="; MU
3371 PRINT "SIGMA="; SIGMA
3372 GOTO 3377
3373 INPUT "MU=? ", MU
3374 INPUT "SIGMA=? ", SIGMA
3377 FOR I=1 TO 27
3378 X=S(I)*SIGMA+MU
3379 IF AA=0 GOTO 3385
3380 X=EXP(X)
3385 PRINT PS(I); " "; X
3387 NEXT I
3390 CLS
3391 CURSOR 6
```

```
3392 PAUSE "END"
3394 GCURSOR 55
3395 GPRINT "02 3A 2F 3A 42 3F 55 7F"
3396 END
3400 I0 = -4 : I1 = 4 : I2 = 0
3410 FOR J = 0 TO 5
3420 FOR I = I0 TO I1
3430 Z = (I2 + I) / 10 ^ J
3440 GOSUB 4000
3450 IF FU >= FX GOTO 3470
3460 NEXT I
3470 I2 = (I2 + I - 1) * 10
3480 I0 = 0
3490 I1 = 10
3500 NEXT J
3510 U = Z
3520 RETURN
4000 Y = EXP((-1) * Z * Z / 2.0) * 0.39894228
4010 Q = 1.0 / (1.0 + 0.2316419 * ABS(Z))
4020 SUM = 0
4030 FOR K = 1 TO 5
4040 SUM = SUM + H(K) * Q ^ K
4050 NEXT K
4060 ZZ = 1
4070 IF Z > 0 GOTO 4090
4080 ZZ = -1
4090 ERF = 0.5 + ZZ * (0.5 - Y * SUM)
4100 FU = ERF
4105 WAIT
4110 PRINT "      Z = ", Z
4200 RETURN
```

五、应用实例

输入: AA = 0

BB = 1

X1 = 0.41

PX1 = .6591

X2 = 1.1

PX2 = .1357

输出： $\text{MU} = 0.0000407325231$

$\text{SIGMA} = 1.000099348$

P (X)	X	P (X)	X
.99	- 2.3265	.45	0.1257
.98	- 2.0539	.40	0.2534
.96	- 1.7508	.35	0.3854
.94	- 1.5549	.30	0.5245
.92	- 1.4052	.25	0.6746
.90	- 1.2816	.20	0.8417
.85	- 1.0365	.15	1.0366
.80	- 0.8417	.10	1.2817
.75	- 0.6745	.08	1.4050
.70	- 0.5244	.06	1.5549
.65	- 0.3853	.04	1.7509
.60	- 0.2533	.02	2.0540
.55	- 0.1256	.005	2.5761
.50	- 0.00047		

(注：本例是标准正态分布，输入数据 X_1 、 PX_1 、 X_2 、 PX_2 是从正态分布表上查得的，本程序再现了这个分布。)

有时，据前人的经验知道某个参数是正态分布或对数正态分布，根据不同的资料估计了这个分布的两个以上的点，这些点很可能是不协调的，因为每两个点就决定了一个分布，这些点可能分别构成几个不同的分布。本程序可以帮助分析这些点的不协调情况，调整好估计值，使它们更合理。

六、讨 论

本程序的计算精度由语句3410决定，计算准确到小数点后5位到6位。如需再提高，可将程序中由语句3073—3076给出的初值重新赋值即可。

(收稿日期：1988年4月27日)

参 考 文 献

1. 中国科学院数学研究所概率统计室，1974，常用数理统计表，科学出版社。
2. E. C. Capen, The Difficulty of Assessing Uncertainty, Journal of Petroleum Technology, Aug. 1976.

A WAY TO INCREASE THE RELIABILITY OF ESTIMATION

Hu Guiming

(Research Institute of Petroleum Geology, Ministry of
Geology and Mineral Resources)

Abstract

One of the basic problems of quantification in geoscience is uncertainty. Based on the concept of probability, the possible value for a variable is proposed with the estimation of a single value replaced by probability distribution. The reliability of estimating parameters is further discussed.

There are various methods for estimating the reliability of a parameter with probability distribution. Applying the mechanism of feedback, the proposed method of feedback combines organically the precision of mathematic computation and the human judgement. Sharp PC - 1500 computer is used as the working computer for the study.