We discuss a petroleum discovery model that greatly simplifies the approach initiated by Barouch and Kaufman (1976) in which exploration is viewed as a sampling without replacement process, and the probability of discovery of a pool is proportional to its size. Calculations that formerly required lengthy Monte Carlo simulations have been reduced to compact formulas.

Key words:
Exploration
Crude oil
Natural gas
Introduction

In this article, we derive a petroleum discovery model that is based on the approach initiated by Arps and Roberts (1958) and refined by Barouch and Kaufman (1976) and others—namely, sampling without replacement, in which the probability of discovery of a pool of a particular size is proportional to the size and to the numbers of pools of that size. Arps and Roberts (1958) were able to derive a compact formula for the density of remaining pools as a function of pool size and number of exploratory wells drilled, but they assumed that the total area under consideration for new wells did not change appreciably as a result of discoveries. However, the added realism gained by reversing this assumption (for example, Barouch and Kaufman, 1976; Eckbo and others, 1978; Smith, 1980; Power, 1990) came apparently at the expense of losing a compact formula; evaluation was through lengthy Monte Carlo simulation. With the help of an approximation developed by Manly (1974), we have obtained compact formulas for the expected value of the density of remaining pools and for other quantities of interest in a way that incorporates the more realistic approach.

The Basic Model and Manly’s Approximation

Suppose that there are $N$ pools altogether, in $K$ classes that are distinguished by size, with $A_i$ pools in the $i$th class. The size of each pool in the $i$th class is $A_i$. By “size” of a pool, we may mean area, area raised to a positive power, or volume; different researchers have adopted different definitions. The model asserts that the probability that the first discovery is in the $i$th class is

$$\frac{\beta_i A_i}{\sum_{i=1}^{K} \beta_i A_i}.$$ 

After $j$ discoveries have been made, in a realization of the stochastic process, a number, $g_{kj}$, of pools will remain undiscovered in class $k$. The conditional probability that the next discovery (number $j + 1$) will be in the $k$th class, given that the state after discovery $j$ is $g_j = (g_{0j}, g_{1j}, \ldots, g_{Kj})$, is

$$\frac{\beta_k g_{kj}}{\sum_{i=1}^{K} \beta_i g_{ij}}.$$ 

With equation 1, it is possible to employ Monte Carlo simulation to produce mean value forecasts of amounts discovered, as a function of the number of discoveries, and variability measures (for example, Smith, 1980).

Manly (1974) derived difference equations whose solutions approximate the means and variance-covariance matrices of the $g_j$. We shall discuss only the mean’s approximation, that is,

$$\mu_{kj} = \mu_{kj-1} - \beta_k \mu_{kj-1} \sum_{i=1}^{K} \beta_i \mu_{ij-1},$$ 

where $\mu_{kj}$ is the approximate mean value of $g_{kj}$. The difference equation 2 can be understood intuitively, since $\beta_k \mu_{kj-1} / \sum_{i=1}^{K} \beta_i \mu_{ij-1}$ is the expected value of the number of pools found in class $k$ on the next discovery, conditional on the numbers remaining in all classes being equal to their expected values after $j - 1$ discoveries. The difference equation 2 appears to be very accurate (Manly and others, 1972; Fuller and others, 1991; Ninpong and others, 1992).

The Differential Equation

The difference equation 2 can provide an excellent approximation to the mean values in much less time than simulation (Fuller and others, 1991; Ninpong and others, 1992). However, it would be nice to simplify the theory even further, in such a way that the mean values are represented as formulas involving the model’s parameters and the discovery number. With this aim in mind, we have used equation 2 as the motivation for a model in which the discovery number is a continuous variable, and the notion of numbers of pools remaining in a finite number of discrete size classes is replaced by a smooth density function of a continuous pool size variable.

The primary justification for considering a model with continuous variables and smooth functions is to be able to use the tools of calculus. Secondary justifications include the facts that geologists often use smooth density functions of a continuous size variable (for example, the lognormal) to describe the frequencies of occurrence of deposits of various sizes, and that other researchers have devised theories that use functions of a continuous discovery number (for example, Arps and Roberts, 1958; Lucki and Szkutnik, 1989).

In the differential equation developed here, the discovery number $j$ from equation 2 is replaced by the continuous variable $w$, and the class indices $i$ and $k$ are replaced by the continuous pool volume variables $r$ and $s$. We further substitute the function $\beta(s)$ for the quantities $\beta_i$ in equation 2.

The basic idea of the density function is illustrated in figure 1. The height of any block in the histogram gives the number of pools whose sizes fall in the range defined by the left and right edges of the block. The histogram represents the values $\mu_{kj}$, or, for $j = 0$, $A_i$. Therefore, areas