Change of support problem

Alexey Ya. Chervonenkis

July 2001

1 Stages of exploration

Mineral deposits are usually explored in several stages. We shall be interested in two of these stages: the so called detailed exploration, which is aimed at resource estimation and localisation sufficient for annual and possibly monthly scheduling of mining; and exploitation exploration aimed to final localisation of ore and waste volumes. Ore should be further directed to a processing plant and waste is either left underground or transported to waste piles. We shall call these stages simply the first stage and the second stage. Usually, the first stage exploration net density is not sufficient for precise localisation of waste and ore. So, in some sense the first stage is aimed at predicting, on average, the results of the second stage of exploration.

2 Selectivity

Waste and ore volumes are distinguished by their average grade in comparison with a fixed cut-off grade. If the grade exceeds the cut-off the volume is considered as ore, otherwise as waste. But really a grade field in space has a fractal character. If a volume has average grade c, it is quite possible that one half of it has the grade of c/2 and the other has the grade of 2c. The smaller the volume size the greater the grade changibility in space. At the final extreme, corns of useful mineral have a very high grade, and are surrounded by waste rock. For practical mining, the volume of selectivity units is determined by mining technology (it is impossible to mine a deposit by the spoonful) or sometimes by the properties of the surrounding rock. Each such unit is to be considered in whole as ore, or assumed as waste. The problem is that that the unit size is almost always much greater than the size of the exploration samples used to measure the grades. For instance, a mining selectivity unit can be 10x10x10 metres in size, while a core sample is a cylinder of 1 meter length and 15 cm diameter. This factor does not form severe difficulty at the second stage of exploration, but becomes a large problem at the first stage. It is called the Change of support problem.

If we consider the grade field as a random field, then at the second stage of exploration we have to estimate conditional expectation of the field in every point (or at a sufficiently dense net of points) under condition of known grades at sampling points, and to integrate (or average) this value over each selectivity unit. After that, the estimated average grade of a unit is compared with the cut-off grade and the final decision on the unit classification (ore or waste) is taken.

But at the first stage of exploration the sampling net is very poor in comparison with that of the second stage, and it is reasonable not to estimate conditional expectation, but rather to estimate conditional distribution at all points. It will predict possible values (and their probabilities) of the second stage sampling. It would be possible to compare this conditional distribution with the cut-off grade and then calculate the probability of the point grade being higher than cut-off grade. This probability could be assumed as the probability of ore at this point. Averaging this probability over the whole deposit or its part we could obtain an estimate of reserves (resources).

Still this estimate would correspond to the case when the size of selectivity mining unit is equal to the size of samples, which is not the case in practice. The field of average selectivity unit grades is much more smooth than that of samples. It corresponds to distribution compression to the mean value (variance decrease), though in most cases it is not quite affine compression. This distribution transformation due to difference in size of selectivity units and samples is called change of support effect.

But this is not the whole story. Practical classification of the selectivity units is based not on their real average grades, but on the estimations of these grades, calculated on the basis of the second stage exploration sampling. These estimates are still more smooth (in space) than the real values. It in term causes further compression of conditional distribution estimates. This effect is called the information effect. Both effects decrease selectivity and thus decrease economic effectiveness. But to decrease the impact of these effects, one would have to use mining technology capable of extracting a finer unit, which is even more expensive, or/and use more dense sampling at the second exploration stage, which also increases costs. Below we shall consider both effects in detail.

3 Estimation of conditional expectation

The most popular procedure to estimate conditional expectation is krieging. In the case of normal random field, when multidimensional distribution of grades at an arbitrary (finite) set of points is assumed Gaussian, this procedure gives real conditional expectation. It is assumed that expectation $M(x)$ (unconditional), variance $D(x)$ and correlation function $K(x, y)$ are given. Here x and y denote points in space. Then given grade values $c_1, c_2, \dots, c_n$ at points $x_1, x_2, \dots, x_n$ the conditional expectation at the point x is calculated as

$$E\frac{c(x)}{c_1, c_2, \ldots, c_n} = M(x) + \Sigma[\lambda_i(c_i - M(x_i)],$$

where the sum is taken over $i = 1, n$ and the weights are found by solving the krieging equation system

$$\Sigma\lambda_i K(x_i, x_j) = K(x, x_j); \quad j = 1, n, \text{ the sum over } i = 1, n).$$

In the stationary case $M(x)$ is constant and $K(x, y) = K^*(x - y)$, depends only on the shift vector $v = y - x$.

If the field is not Gaussian the krieging technique gives optimal linear estimation of conditional expectation.

There are many modifications of the krieging technique: ordinary krieging, logarithmic krieging, and other nonlinear methods. They take into account the fact that real grade fields are not Gaussian, the field is not stationary and unconditional expectation (as a function of x) is usually not known. If we are interested not in conditional grade expectation at a single point x, but in the average grade over some volume V, it is enough to substitute to the right part of the krieging equations the mean value of $K(x, x_j)$ averaged over $x \in V$ instead of $K(x, x_j)$.

But krieging is used at the last stage of exploration, when the final decision is to be taken. The calculated estimate of conditional expectation is compared with the cut-off grade and if the expectation exceeds the cut-off grade the volume is considered as ore, otherwise assumed to be waste.

4 Estimation of conditional distribution

At the first stage of exploration our goal is to estimate not the conditional expectation at an arbitrary point x, but to estimate the conditional distribution at this point. In the Gaussian case conditional distribution is also Gaussian, and is totally determined by the conditional expectation and the conditional variance. (The latter is also calculated using krieging technique). But we shall not stay on this case because real grade distributions are far from Gaussian.

But conditional distribution estimates calculated in this way are conditional distributions of grades in the units of sample size. If we average the right side of krieging equations, as we did before, we obtain conditional distribution of point grades averaged over the volume V, but not conditional distribution of grades averaged over the volume.

5 Change of support problem

Conditional distribution change due to change of support and information effects

The change of support and information effects can be combined in one formal scheme. Let $P(c)$ be conditional distribution density of the grade at a point x under condition of known grades at points $x_1, x_2, \ldots, x_n$ as the result of the first stage exploration sampling. At the second stage we get additional information, which we denote by some abstract vector q. After this the conditional distribution density is changed to $P^*(c, q)$, where q is a parameter. We assume that there is some distribution density $p(q)$ (also conditional) of possible q-values. As far as in average all $P^*(c, q)$ give $P(c)$ we have the following integral relation

$$P(c) = \int p(q) P^*(c, q) dq$$

Instead of q we can use the resulting (after the second stage of exploration) grade estimate c_v of the selectivity unit grade surrounding the point x. Then we have

$$P(c) = \int p(c_v) P^*(c, c_v) dc_v \tag{1}$$

where c_v is used as a parameter for the value c distribution density $P^*(c, c_v)$. Of course in this case

$$\begin{aligned} &\textstyle\int P*(c, c_v) dc = 1, \\ &\textstyle\int cP*(c, c_v) dc = c_v. \end{aligned}$$

So the probability density $P(c)$ is the result of smoothing the density $p(c_v)$ by the kernel $P^*(c, c_v)$.

Integral equation

The kernel $P^*(c, c_v)$ is the conditional distribution density of point values at the point x. Formally it depends (as a parameter) not only on c_v but also on the result of first stage sampling. But really in many cases this influence is only slight, and the kernel can be considered as constant at least for some parts of a deposit. We shall discuss the problem of its estimation later. Now we can consider the expression (1) as an integral equation for unknown function $p(c_v)$ with known right part (estimated as the result of first stage sampling) and known kernel $P^*(c, c_v)$.

As far as we have only estimations of conditional distributions $P(c)$ and $P^*(c, c_v)$, it is reasonable to use approximation to this equation in the form of a (continuous) linear programming problem: find $p(c_v)$ such that

$$
\begin{array}{ll}
a) & \int p(c_v)dc_v = 1, \\
b) & \int c_v p(c_v)dc_v = \int cP(c)dc \\
c) & \int |P(c) - \int p(c_v)P^*(c, c_v)dc_v|dc = \min
\end{array}
$$

Of course for calculations it is necessary to transform the problem to a discrete form.

Approximation by linear programming

For this purpose we change continues densities to discrete histograms. We divide the range of possible c values to k intervals $(a_0 <= c < a_1), (a_1 <= c < a_2), \ldots, (a_{k-1} <= c < a_k)$ and define

$$P_i = \int_{a_{i-1},a_i} P(c)dc, p_j = \int_{a_{j-1},a_j} p(c_v)dc_v, P^*_{i,j} = \int_{a_{i-1},a_i} P(c, c_v)dc(a_{j-1} <= c_v < a_j).$$

We assume that the values P_i and $P^*_{i,j}$ are known, while p_j values are to be found.

Now the problem may be formulated as follows: Find p_j $(1 <= j <= k)$ such that

$$
\begin{array}{l}
0 <= p_j, \\
\Sigma_j p_j = 1, \\
\Sigma_i |\Sigma_j p_j P^*_{i,j} - P_i| \to \min.
\end{array}
$$

The resulting histogram presents the estimation of conditional grades distribution in a selective unit, surrounding the point x.

If the number of intervals is not large (less than around fifty) the linear programming problem is not hard. We can add also conditions concerning mean grade, supposing that conditional mean $C(x)$ of the initial histogram is P_i known. It is

$$\Sigma_j c^*_j p_j = C(x),$$

where c^*_j are either just the middles of corresponding intervals, or average grade of the initial histogram over j-th interval.

Gaussian case

In the Gaussian case everything is much more simple. Let us assume that $P(c) \div N(C(x), D(x)), p(c_v) \div (C(x), D_v(x)), P^*(c, c_v) \div N(c_v, d_v(x))$. Here is assumed that, the values $C(x), D(x)$ and $D_v(x)$ are known. Then it is easily seen that equation (1) is satisfied if

$$D_v(x) = D(x) - d_v(x)$$

So average grade conditional distribution in the selectivity unit surrounding point x is

$$N(C(x), D(x) - d_v(x)).$$

The value $d_v(x)$ may be chosen to be either constant or proportional to $D(x)$. The case of constant variance d_v corresponds to the case of a stationary Gaussian field and a rather thick and uniform sampling at the last stage of exploration. In this case the variance $D(x)$ reflect the closeness of the point x to the points of sampling (at the first stage of exploration). The case of $d_v(x)$ proportional to $D(x)$ corresponds to the state when the real field is not stationary and difference of local variances is large. But in this case the very assumption of the field being Gaussian seems doubtful.

How can we estimate the kernel?

1. **First case:** We have enough data on the second stage of exploration sampling.

In this case, for each area we can both estimate conditional expectation of each selectivity unit (using data krieging), and estimate conditional point distribution averaged over the unit. The mean value of the distribution should coincide with the conditional expectation. The conditional expectation is used as the estimate of the average unit grade. Now we can group the units by their average grade in the intervals of a histogram, and average corresponding conditional densities $P(c)$. So we get c distribution densities $P^*(c, c_v)$ for every interval $a_{j-1} <= c_v < a_j$. If for some intervals there are no data or not enough data, the distribution densities may be obtained by interpolation from the neighbouring ones. Now discretising each of these distribution densities to intervals $a_{i-1} <= c < a_i$ we obtain the kernel matrix P^*_{ij}.

2. **Second case:** no (or not enough) data on the second stage of exploration.

In this case we can use only data presented as the result of the first stage exploration sampling. Still, we are able (or imagine that we are able) to estimate total point variance D and averaged (point) correlation function $K(x, y)$ (or variogramm). It is possible, of course, to first use some transformation of the data and then estimate the variance and correlation function of the transformed data. It is possible to calculate the variance of selectivity unit average grades in the stationary case and uniform selectivity units. (The same is possible for transformed data and averaged over units transformed grades.)

It is

$$\mathrm{Var}(c_v) = D_v = \frac{D}{|v|^2} \int_v \int_v K(x - y)\, dx\, dy \tag{2}$$

So the ratio $R = \frac{D_v}{D}$ is equal to $|v| - 2 \int_v \int_v K(x - y)\, dx\, dy$ and may be either computed or calculated analytically.

a) Gaussian kernal for initial data

First we consider the case when no prior data transformation is made. Imagine that the second stage of exploration has a rather thick sampling net and that the selectivity unit size is comparable to or less than correlation radius of the point sampling data. Then we can suppose that conditional standard deviation of the point data within a selectivity unit under the conditions of a given unit grade estimate is small in comparison with the average grade of the unit. In this case the density $P^*(c, c_v)$ may usually be approximated by the Gaussian distribution $N(c_v, D(c_v))$. Then the only problem is to find the variance dependency $D(c_v)$ on the mean value c_v. Usually for geological sampling data (mulistructural model) the grade standard deviation is proportional to the expectation. So we can assume that $D(c_v) = k(c_v)^2$, with some unknown coefficient k. Really this coefficient depends on natural grade changibility within a selectivity unit and on the relation between the unit size and a sample size and form. So the problem is reduced to k estimation.

Now, the integral relation (1) is valid not only for local distribution densities but also for total ones. So for any fixed k we can substitute the proposed kernel, and obtain the distribution density of v units average grades. Having this distribution we can find its variance and change k until the variance stisfies the equation (2). These calculations can be produced once for the whole deposit or independently for several large divisions of it. As the result we obtain the kernel.

b) Non-Gaussian kernal for initial data

The same procedure can be carried out for any other type of kernel (corresponding to conditional grade distribution $P^*(c, c_v)$), which is characterised by two parameters, determining the distribution expectation and variance. It could be gamma distribution, lognormal or some other. Again choosing one parametrical dependency between the expectation and the variance of the distribution $P^*(c, c_v)$ we can adjust this parameter, solving integral equation (1) (for the total distribution) and fitting the relation (2).

c) Preliminary data transformation

To simplify the procedure and to use more reasonable types of kernels, preliminary data transformation can be used. Usually one looks for data transformations that transform the total distribution to a given standard type. In could be Gaussian distribution, gamma distribution or some other. Usually it is assumed that the distribution $P^*(c, c_v)$ is also transformed to the same type. Now we can carry out all the above procedure for the transformed data (including R ratio calculation). As the result we obtain $P^*(t, t_v)$, where t denotes transformed grade and t_v denotes its average value over a unit. It is more resonable to use here a constant variance of the distribution $P^*(t, t_v)$ (and look for its value fitting relation (2) for transformed data) because the transformation usually eliminates dependency between variance and expectation. Solving the integral equation (1) usually becomes trivial (as was shown above for Gaussian case). Then, by back transformation, the kernel $P^*(c, t_v)$ can be found. It is important to notice that c_v is not equal to the back transformaton of t_v. But it can be easily found as the expectation of the distribution $P^*(c, t_v)$.

Conclusion

So we have proposed several ways to find the decision of the Change of support problem, including change of support and information effects. All of them are based on the idea of approximately solving the integral equation (1) under restrictions making the result consistent. It is important to notice that it is not supposed that conditional distribution $P(c)$ is of a certain parametrical type, nor that data transformation reduces the total distribution to a fixed type does the same for conditional distributions. Still any technique for average unit grade distribution estimation based on conditional point distribution has the following drawback. The data used for estimation of the conditional point distribution are sampled at different distances from the point. This information is lost in the distribution, though really it is available. There are other methods taking in account this fact. But it is another story.