МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Томский государственный университет
Горно-Алтайский государственный университет
Институт оптики атмосферы им. В.Е. Зуева СО РАН

НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ИССЛЕДОВАНИИ СЛОЖНЫХ СТРУКТУР

МАТЕРИАЛЫ ДЕСЯТОЙ РОССИЙСКОЙ КОНФЕРЕНЦИИ С МЕЖДУНАРОДНЫМ УЧАСТИЕМ

Томск
Издательский Дом Томского государственного университета
2014

С помощью модели накопления было исследовано распределение численности регионов РФ. Проведенные исследования можно использовать для анализа и прогноза изменения численности населения, проследить динамику численности и плотности населения.

Литература

1. *Дмитриев Ю.Г., Устинов Ю.К.* Математические модели растущих систем // Вычислительные технологии. 2001. Т. 12. С. 68–75.
2. *Талейко А.В.* Закономерности формирования численности населения по территориям // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2009. № 3.
3. *Федеральная* служба государственной статистики. М., 1999–2014. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/ (дата обращения: 11.04.2014).

КОМБИНИРОВАННАЯ ОЦЕНКА ЛИНЕЙНОГО ФУНКЦИОНАЛА

Ю.Г. Дмитриев

Национальный исследовательский Томский государственный университет, Томск, Россия
dmit@mail.tsu.ru

Пусть $X_1, \ldots, X_n$ – независимые наблюдения объема n над случайной величиной X с неизвестной функцией распределения $F \in \boldsymbol{F}$. Требуется оценить линейный функционал

$$J(F) = M_F[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) dF(x), \tag{1}$$

здесь φ – заданная вещественная функция. Непараметрической оценкой данного функционала является $\hat{J} = J(F_n)$, где $F_n(t)=n^{-1} \sum_{i=1}^{n} c(t-X_i)$, $c(t)=\{0: t<0, 1: t\geq 0\}$ – эмпирическая функция распределения (э.ф.р.).

Пусть имеется предположение, что $F \in \boldsymbol{G} = \{G(x;\theta),\ \theta \in \Theta\}$. В этом случае (1) запишется в виде $J(G) = \Psi(\theta)$.

Величину $\Psi(\theta)$ назовем априорной догадкой, она выступает в качестве априорной оценки (возможного значения) функционала J. Задача состоит в построении оценки функционала (1), учитывающей совместно непараметрическую оценку и априорную догадку. Следуя работе [1], рассмотрим комбинированную оценку вида

$$\hat{J}_\lambda = (1-\lambda)\hat{J} + \lambda\Psi(\theta) = \hat{J} - \lambda(\hat{J} - \Psi(\theta)), \tag{2}$$

где весовой коэффициент λ, выбранный из условия минимума среднеквадратической ошибки (СКО) $S_F^2(\lambda) = M_F[\hat{J}_\lambda - J]^2$, определяется выражением

$$\lambda = \lambda(F) = \frac{\sigma_F^2}{\sigma_F^2 + n\Delta_F^2} = \left(1 + n\frac{\Delta_F^2}{\sigma_F^2}\right)^{-1}. \tag{3}$$

Здесь $\sigma^2{}_F = D_F[\varphi(X)]$ – дисперсия $\varphi(X)$, а $\Delta_F = J(F) - \Psi(\theta)$ величина смещения (отклонения) априорной догадки от истинного значения $J(F)$.

Минимум СКО удовлетворяет соотношению $nS_F^2(\lambda) = \sigma_F^2(1-\lambda)$. Весовой коэффициент λ изменятся в пределах $0< \lambda \leq 1$ и показывает, какое влияние оказывает каждая из оценок в комбинированной оценке (2). При $\Delta_F = 0$ имеем $\lambda=1$ и в качестве оценки функционала $J(F)$ следует взять априорную догадку $\Psi(\theta)$. При $\Delta_F \neq 0$, что, обычно, и бывает на практике, $\lambda<1$ и с ростом объема наблюдений ($n\to\infty$) $\lambda\to 0$, влияние априорной догадки уменьшается, выигрыш в точности оценивания убывает.

Коэффициент λ зависит от неизвестной функции распределения F и поэтому его значение, как правило, неизвестно. Пусть значение θ задано. Подставим в формулу (3) вместо F э.ф.р. F_n. Тогда получим оценку

$$\hat{\lambda} = \lambda(F_n) = \left(1 + n\hat{\Delta}^2 / \hat{\sigma}^2\right)^{-1}.$$

Подставляя $\hat{\lambda}$ в (2), получаем адаптивную комбинированную оценку $\tilde{J} = \hat{J} - \hat{\lambda}(\hat{J} - \Psi(\theta))$. Рассмотрим асимптотическое поведение $\tilde{J}$. Обозначим

$$\xi_n = \frac{\sqrt{n}(\tilde{J} - J)}{\hat{\sigma}} = \eta_n - \frac{\eta_n + b_n}{1 + (\eta_n + b_n)^2}, \text{ где } \eta_n = \sqrt{n}(\hat{J} - J)\big/\hat{\sigma},\ b_n = \sqrt{n}\Delta_F\big/\hat{\sigma}.$$

Пусть при $n \to \infty$ последовательность b_n стремится по вероятности к b. Тогда в силу центральной предельной теоремы и теоремы непрерывности ([2], глава 6) случайная последовательность ξ_n стремится по распределению к случайной величине $\xi = \eta - (\eta + b)/(1 + (\eta + b)^2)$, где $\eta \in N(0,1)$ – стандартная нормальная случайная величина.

При $\Delta_F = 0$, $b = 0$ математическое ожидание $M\xi = 0$ и дисперсия $D\xi = 0{,}4670388$. При $b = \pm\infty$ случайная величина $\xi = \eta$ с $M\xi = 0$ и дисперсией $D\xi = 1$. В этом случае асимптотические распределения адаптивной оценки $\tilde{J}$ и непараметрической оценки $\hat{J}$ одинаковы и совпадают с нормальным законом $N(J,\sigma^2_F)$. Поскольку функция $y = x^3/(1 + x^2)$ монотонно возрастает на числовой прямой, распределение величины ξ легко расчитывается для заданных значений b с помощью стандартного нормального закона.

Литература

1. *Дмитриев Ю.Г., Скрипин С.В.* О комбинированной оценке вероятности безотказной работы по полной выборке // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 4. С. 32–38.
2. *Боровков А.А.* Математическая статистика. М. : Наука, 2007. 704 с.

ОЦЕНКА ЗАПАСОВ РУД ПО ДАННЫМ ОПРОБОВАНИЯ НА СТАДИИ ПОИСКА

Ю.Г. Дмитриев, Т.О. Кошевая

Национальный исследовательский Томский государственный университет, Томск, Россия
dmit@mail.tsu.ru, dodowoman@mail.ru

Одним из этапов геологоразведочных работ является поиск полезных ископаемых. Практика показывает [1], что из числа месторождений, подвергаемых предварительной разведке, промышленными оказываются лишь 5–10%. Такая низкая результативность объясняется несовершенством методов прогнозирования (оценивания) запасов полезных ископаемых на стадии поисковых работ. Основная причина этого состоит в том, что из-за редкой сети опробования не удается достаточно хорошо оконтурить рудные тела, рассчитать их объем и среднее содержание полезного компонента. Величины запасов рассчитываются по формуле

$$Q = \frac{V\bar{C}d}{100\%},$$

где Q – величина запаса полезного компонента, т; V – объем промышленных руд, м^2; $\bar{C}$ – среднее содержание полезного компонента в рудах, выраженное в процентах и полученное на основании результатов опробования; d – объемный вес руды, т/м^2. Располагая значениями и частотой встречаемости различных концентраций полезного ископаемого, можно построить гистограмму распределения содержаний в опробованном объеме рудоносных пород исследуемого рудоносного участка. Затем, используя статистическую зависимость между средними содержаниями и соответствующими долями объемов опробованных пород, можно спрогнозировать (оценить) величину запаса полезного компонента для заданного уровня промышленного содержания. Такая методология изложена в работе [1]. В докладе развиваются идеи этой работы в различных направлениях. Наряду с гистограммой для анализа данных применяется аппроксимация бета – распределением [2]. Прогнозирование осуществляется с учетом функциональной и структурной зависимости между переменными [3], а также модели накопления вещества для растущих систем [4]. Вид данных, по которым проводились исследования, представлен в таблице:

В таблице представлены результаты опробования скважин, пробуренных на редкометальном месторождении по сети 400x400 м.Общий объем опробований равен 419. В столбцах 1-3 приведены результаты концентраций по данным опробованиям и частота их появления, В 4 столбце даны средние концентрации, в 5 и 6 средние концентрации и доли объема в усечённых эмпирических распределениях, C=10,8.

Для различных вариантов усечений исходного эмпирического распределения проанализированы зависимости между полученными концентрациями (столбец 7) и долями объема (столбец 6).