

УДК 553.048

МРНТИ 38.01.77

DOI: <https://doi.org/10.54859/kjogi108590>

Получена: 05.09.2022.

Одобрена: 01.12.2022.

Опубликована: 27.12.2022.

Оригинальное исследование

Подбор оптимального числа симуляций Монте-Карло при вероятностном подсчёте ресурсов углеводородного сырья

Р.М. Садыков

КМГ Инжиниринг, г. Астана, Казахстан

АННОТАЦИЯ

Обоснование. Вероятностный подсчёт ресурсов используется при оценке потенциала нефтегазового проспекта, что является основой при принятии решений в нефтегазовой индустрии. Основным результатом вероятностного подсчёта ресурсов является оценка ресурсного потенциала, выраженная в виде вероятностной функции распределения геологических или извлекаемых ресурсов углеводородов. При этом часто используется симуляция Монте-Карло. Она позволяет находить численное значение целевой функции, в случае с вероятностным подсчётом – вероятностной функции распределения запасов.

Цель. Данная работа нацелена на нахождение оптимального количества симуляций при вероятностном подсчёте ресурсов.

Методы. Симуляция Монте-Карло основана на многократном повторении расчёта функции, описывающей процесс, с использованием генератора случайных чисел. Переменная функции, задаваемая одним из статистических распределений, рассчитывается по этому случайному числу. Далее над всеми переменными функции проводятся математические операции согласно математической модели. Обобщая полученное множество результатов, приближенно рассчитывается статистическое распределение, описывающее искомую функцию.

Результаты. Точность определения функции распределения растёт при увеличении числа симуляций. Однако это приводит к увеличению времени вычисления. Таким образом, возникает выбор между скоростью и точностью решения.

Латинский гиперкуб позволяет уменьшить влияние генератора случайных чисел, но дополнительные промежуточные вычисления нивелируют уменьшение числа симуляций при использовании латинского гиперкуба по сравнению со случайной выборкой.

Заключение. По результатам работы была получена логарифмическая зависимость точности расчетов от числа симуляций Монте-Карло. В большинстве случаев точность результатов в 1% при 10000 симуляций будет достаточной.

Ключевые слова: статистика, симуляция Монте-Карло, латинский гиперкуб, имитационное моделирование, вероятностный подсчёт запасов.

Как цитировать:

Садыков Р.М. Подбор оптимального числа симуляций Монте-Карло при вероятностном подсчёте ресурсов углеводородного сырья // *Вестник нефтегазовой отрасли Казахстана*. 2022. Том 4, №4. С. 32–44. DOI: <https://doi.org/10.54859/kjogi108590>

UDC 553.048

CSCSTI 38.01.77

DOI: <https://doi.org/10.54859/kjogi108590>

Received: 05.09.2022.

Accepted: 01.12.2022.

Published: 27.12.2022.

Original article

Optimal number of Monte-Carlo simulations for hydrocarbon resources probabilistic estimation

Raman M. Sadykov

KMG Engineering, Astana, Kazakhstan

ABSTRACT

Background: Probabilistic resource estimation allows to assess the potential of oil and gas prospect, which is used for decision-making in the industry. The main result of the estimation is resource potential, expressed in probability distribution function of geological or recoverable hydrocarbon resources. Monte Carlo simulation commonly used for these purposes. It simulates the numerical value of the objective function.

Aim: The aim of the work is to find the optimal number of Monte-Carlo simulations in the probabilistic resource estimation.

Methods: Monte Carlo simulation is based on repeated calculation of the function describing the process using a random number generator. The function variable defined by statistical distributions is calculated using random number. Further, mathematical operations are performed on all variables of the function according to the mathematical model. Summarizing the obtained set of results, the statistical distribution describing the desired function is approximately estimated.

Results: The accuracy of distribution function estimation increases with an increase of the number of simulations. However, this tend to increase the computation time. Thus, there is a choice between the speed and accuracy of the solution.

The Latin hypercube reduces the influence of the random number generator imperfection, but intermediate calculations with Latin hypercube attenuate the decrease in the number of simulations compared to random sampling.

Conclusion: Based on the results of the work, a logarithmic dependence of the calculation accuracy on the number of Monte Carlo simulations was obtained. In most cases, 10000 simulations with accuracy about 1% will be sufficient.

Keywords: *statistics, Monte Carlo simulation, Latin hypercube, simulation modeling, probabilistic reserve calculation.*

To cite this article:

Sadykov RM. Optimal number of Monte-Carlo simulations for hydrocarbon resources probabilistic estimation. *Kazakhstan journal for oil & gas industry*. 2022;4(4):32–44. DOI: <https://doi.org/10.54859/kjogi108590>

ӨОЖ 553.048

ГТАХР 38.01.77

DOI: <https://doi.org/10.54859/kjogi108590>

Қабылданды: 05.09.2022.

Мақұлданды: 01.12.2022.

Жарияланды: 27.12.2022.

Түпнұсқа зерттеу

Көмірсутек шикізаты ресурстарының ықтималдығын есептеу кезінде Монте-Карло симуляцияларының оңтайлы санын таңдау

Р.М. Садықов

ҚМГ Инжиниринг, Астана қаласы, Қазақстан

АННОТАЦИЯ

Негіздеме. Ресурстарды ықтималды есептеу мұнай-газ индустриясында шешім қабылдау кезінде негіз болып табылатын мұнай-газ болашағының әлеуетін бағалау кезінде қолданылады. Ресурстардың ықтималдығын есептеудің негізгі нәтижесі көмірсутектердің геологиялық немесе өндірілетін ресурстарын бөлудің ықтималдық функциясы түрінде көрсетілген ресурстық әлеуетті бағалау болып табылады. Бұл ретте Монте-Карло симуляциясы жиі қолданылады. Бұл мақсатты функцияның сандық мәнін, ықтималдықты есептеу жағдайында – қорларды бөлудің ықтималдық функциясын табуға мүмкіндік береді.

Мақсат. Аталған жұмыс ресурстардың ықтималдығын есептеу кезінде симуляцияның оңтайлы санын табуға бағытталған.

Тәсілдер. Монте-Карло симуляциясы кездейсоқ сандар генераторын қолдана отырып, процесті сипаттайтын функцияны есептеуді бірнеше рет қайталауға негізделген. Статистикалық үлестірімдердің бірімен берілген функцияның айнымалысы осы кездейсоқ саннан есептеледі. Бұдан әрі математикалық модельге сәйкес функцияның барлық айнымалылары бойынша математикалық амалдар орындалады. Алынған нәтижелер жиынтығын қорытындылай келе, қажетті функцияны сипаттайтын статистикалық үлестіру шамамен есептеледі.

Нәтижелер. Үлестіру функциясын анықтау дәлдігі симуляция санының өсуімен артады. Алайда, бұл есептеу уақытының ұлғаюына әкеледі. Осылайша, шешім жылдамдығы мен дәлдігі арасында таңдау пайда болады.

Латын гиперкубы кездейсоқ сандар генераторының әсерін азайтуға мүмкіндік береді, бірақ қосымша аралық есептеулер кездейсоқ іріктеумен салыстырғанда латын гиперкубын қолданған кезде симуляция санының азаюын теңестіреді.

Қорытынды. Жұмыстың нәтижелері бойынша Монте-Карло симуляция санына есептеу дәлдігінің логарифмдік тәуелділігі алынды. Көп жағдайда 10000 симуляция кезінде 1% нәтиже дәлдігі жеткілікті болады.

Негізгі сөздер: статистика, Монте-Карло симуляциясы, латын гиперкубасы, имитациялық модельдеу, қорларды ықтималдық есептеу.

Дәйексөз көлтіру үшін:

Садықов Р.М. Көмірсутек шикізаты ресурстарының ықтималдығын есептеу кезінде Монте-Карло симуляцияларының оңтайлы санын таңдау // Қазақстанның мұнай-газ саласының хабаршысы. 2022. 4 том, №4. 32–44 б. DOI: <https://doi.org/10.54859/kjogi108590>

Введение

Симуляция Монте-Карло – широко используемая группа методов, в которых процесс или функция изучается путём многократных экспериментов (реализаций) с использованием определенной математической модели [1]. Симуляция используется в случаях, когда переменные функции неизвестны, но известно их распределение, а процесс или функция являются достаточно сложными для поиска аналитического решения.

Например, произведение двух случайных величин X , Y даёт третью случайную величину $X \cdot Y$, дисперсия которой описывается достаточно громоздким уравнением. При использовании усечённых распределений определение дисперсии результата значительно усложняется.

$$\text{Var}(X \cdot Y) = \text{Var}(X) \cdot \text{Var}(Y) + E(X)^2 \cdot \text{Var}(Y) + E(Y)^2 \cdot \text{Var}(X) \quad (1)$$

При вероятностном подсчёте запасов объёмным методом используются по крайней мере 6 параметров, следовательно, рассчитать распределение геологических запасов аналитическим уравнением является сложной задачей. Использование усечённых функций распределения в качестве входных данных также усложняет задачу.

Симуляция методом Монте-Карло позволяет рассчитывать процесс с достаточно большим количеством переменных. Основным недостатком метода является приближённость решения. Целью данной работы является определение оптимального числа симуляций при вероятностном подсчёте запасов для получения решения с необходимой погрешностью.

Методология

Симуляция Монте-Карло

Пусть процесс описывается функцией (2), включающей в себя m переменных. В случае с вероятностным подсчётом запасов таковыми параметрами являются: положение контакта, доля коллектора, коэффициент пористости, коэффициент насыщенности углеводородами, пересчётный коэффициент и пр.

$$y = f(X_1, X_2, \dots, X_m) \quad (2)$$

Каждая переменная может быть представлена в виде случайной величины X_i , которая описывается плотностной функцией вероятности $PDF_i(X_i)$ и вероятностной

функцией распределения $CDF_i(X_i)$ [2]. При этом значение случайной величины при определённой вероятности p рассчитывается обратной функцией от функции распределения:

$$X_i = CDF_i^{-1}(p) \quad (3)$$

При расчёте реализации для каждой случайной величины генерируется случайное число, в результате получается m случайных чисел $rand$. В таком случае уравнение (2) приобретает вид (4):

$$y = f\left(CDF_1^{-1}(rand_1), CDF_2^{-1}(rand_2), \dots, CDF_m^{-1}(rand_m)\right) \quad (4)$$

Далее реализации повторяются n раз, и при описании результатов указывается, что было проведено n симуляций.

Таким образом, по результатам симуляции Монте-Карло получают массив результатов $\{y_1, y_2, \dots, y_n\}$, являющихся n -выборкой из функции, описывающей процесс. Предположив, что реализации между собой равновероятностны, итоговую функцию можно приближенно описать следующими статистическими величинами (табл. 1).

Таблица 1. Статистические величины для описания функции распределения
Table 1. Statistic values describing distribution function

Параметр Parameter	Формула Formula
Среднее Mean	$\mu \approx \frac{\sum y_i}{n}$
Среднеквадратичное отклонение Standard deviation	$\sigma \approx \sqrt{\frac{1}{n-1} \sum (y_i - \mu)^2}$
MIN	$min \approx \min\{y_1, y_2, \dots, y_n\}$
MAX	$max \approx \max\{y_1, y_2, \dots, y_n\}$

Латинский гиперкуб и случайная выборка

Случайная выборка (Random sampler) – это способ получения случайной величины, при котором для его расчёта используется случайное число, определенное независимо от предыдущих реализаций. Однако, учитывая, что генератор случайных чисел в большинстве случаев не даёт равномерного распределения, случайная выборка может привести к значительному искажению итоговых результатов. Для пре-

одоления влияния генератора случайных чисел используются разнообразные способы выборки, одним из которых является латинский гиперкуб.

Латинский гиперкуб (Latin-hypercube sampler) – это способ выбора значений из нескольких распределений, который обеспечивает более полный охват диапазона входных данных [3]. Это достигается разделением вертикальной оси накопленной функции распределения на n равных, непересекающихся интервалов (рис. 1). Таким образом, на горизонтальной оси также получается n интервалов с одинаковой

вероятностью наступления, но разной шириной.

Следующим этапом является выбор значений внутри n интервалов. Для этого внутри каждого вертикального интервала случайной выборкой подбирается вероятность p_i (5), которой соответствует некоторое значение внутри интервала на горизонтальной оси (3).

$$p_i = p_{i \min} + rand(p_{i \max} - p_{i \min}) \quad (5)$$

где $p_{i \min}$, $p_{i \max}$ – минимальное и максимальное значения вероятности i -го интервала.

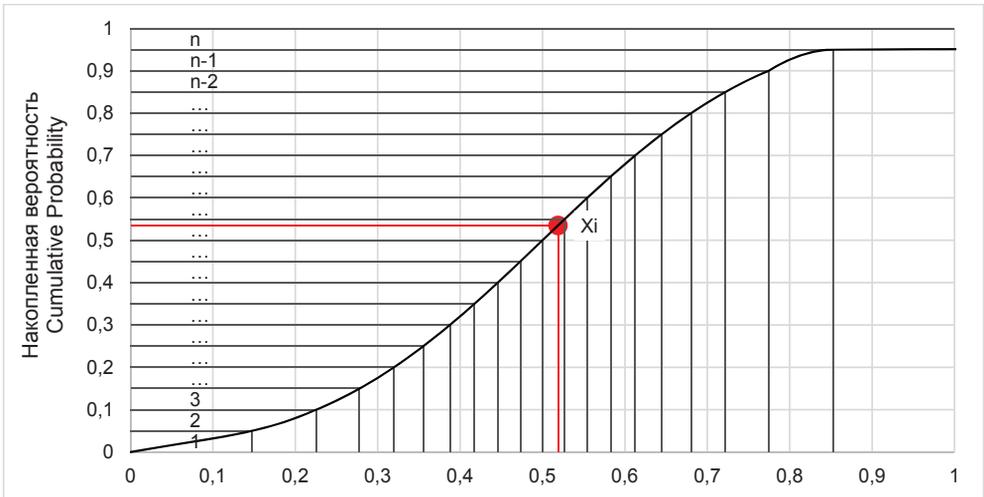


Рисунок 1. Разбиение накопленной функции распределения на n интервалов
Figure 1. Partition of cumulative distribution function into n intervals

Виды распределений

Равномерное распределение – это вид распределения, при котором случайная величина равновероятно может принять любое значение из диапазона от минимального до максимального значения. Для непрерывной случайной вели-

чины функция плотности равномерного распределения есть постоянная. Непрерывное равномерное распределение задается двумя параметрами: минимальное min и максимальное max значения случайной величины [4] (табл. 2).

Таблица 2. Основные виды распределений
Table 2. Main types of distributions

Параметр / Parameter	Равномерное распределение / Uniform distribution	Нормальное распределение / Normal distribution
Плотность распределения / Distribution density	$PDF(x) = \begin{cases} \frac{1}{max - min}, & x \in [min, max] \\ 0, & x \notin [min, max] \end{cases}$	$PDF(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-E_x)^2}{2\sigma^2}}$
Функция распределения / Distribution function	$CDF(x) = \begin{cases} 0, & x < a \\ \frac{x - min}{max - min}, & a \leq x < b \\ 1, & x \geq b \end{cases}$	$CDF(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x - E_x}{\sqrt{2\sigma^2}} \right) \right]$ $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$

Параметр Parameter	Равномерное распределение Uniform distribution	Нормальное распределение Normal distribution
Среднее Mean	$E_x = \frac{\min + \max}{2}$	Исходный параметр Input parameter
Среднеквадратичное отклонение Standard deviation	$\sigma = \sqrt{\frac{(\max - \min)^2}{12}}$	Исходный параметр Input parameter

В случае вероятностного подсчёта ресурсов равномерное распределение используется для определения положения контакта в тех случаях, когда отсутствует какая-либо информация о положении контакта.

Нормальное распределение – наиболее часто используемый вид распределения, в котором функция плотности вероятности задаётся функцией Гаусса (табл. 2). Одной из отличительных особенностей данного распределения является его симметричность относительно среднего значения [5].

При вероятностном подсчёте ресурсов среднюю пористость и насыщенность пород часто описывают с использованием нормального распределения. Кроме того, в процессе вероятностного подсчёта ресурсов используют логнормальное распределение и гамма-распределение.

Отличительной особенностью данных распределений является смещённость плотности распределения в область меньших значений, а также невозможность их использования для случайных величин с отрицательными значениями. При вероятностном подсчёте запасов результирующая величина (геологические или извлекаемые ресурсы) имеет распределение

похожее и на логнормальное, и на гамма-распределение.

Усечение распределения – это преобразование распределения таким образом, что случайная величина не может выпасть за указанные рамки, т.е. случайная величина ограничена минимальным и максимальным значением (6).

$$CDF_{tr}(x) = \frac{CDF(x) - CDF(\min)}{CDF(\max) - CDF(\min)} \quad (6)$$

Вероятностный подсчёт ресурсов Структура X

Структура X представлена брахиантуклиальной складкой и была выявлена по данным 2D сейсмоки. Нефтегазоносность региона подтверждена наличием месторождений поблизости. По данным месторождений-аналогов, а также учитывая глубины, предполагается, что в случае обнаружения залежи пластовым флюидом будет нефть. Подсчётные параметры были определены на основе данных месторождений-аналогов.

Учитывая высокую степень неопределённости, подсчёт ресурсов ведётся вероятностным способом. Исходные данные представлены в табл. 3.

Таблица 3. Исходные данные для вероятностного подсчёта ресурсов
Table 3. Input data for the probabilistic resources estimation

Параметр Parameter	Среднее Mean	Среднеквадратичное отклонение Standard deviation	MIN	MAX
Положение водонефтяного контакта, м Position of the oil/water contact, m			-2360	-2410
Общий объём, тыс. м ³ Total volume, thousand m ³			46018	4064690
Общая толщина, м Total thickness, m	10	3	5	15
Доля коллектора, д. ед. Net to Gross ratio, units	0,7	0,08	0,6	0,8
Коэффициент пористости, д. ед. Porosity, units	0,2	0,03	0,15	0,25
Коэффициент нефтенасыщенности, д. ед. Oil saturation factor, units	0,68	0,1	0,6	0,75

Параметр Parameter	Среднее Mean	Среднеквадратичное отклонение Standard deviation	MIN	MAX
Объёмный коэффициент, д. ед. Formation volume factor, units	0,8	0,05	0,7	0,9
Плотность нефти, т/м ³ Oil density, t/m ³	0,849	0,05	0,7	0,9
Коэффициент извлечения нефти, д. ед. Oil recovery factor, units	0,3	0,05	0,2	0,4
Газосодержание, м ³ /т Gas content, m ³ /t	126,9	30	50	150

Объёмный метод подсчёта ресурсов / запасов

При оценке ресурсного потенциала структуры был использован объёмный метод подсчёта ресурсов нефти. При расчёте геологических m_{geo} и извлекаемых m_{rec} ресурсов нефти производится многократное произведение общего объёма залежи BV на различные коэффициенты (7–8). Геологические VG_{geo} и извлекаемые VG_{rec} ресурсы попутного газа также рассчитываются путём перемножения соответствующих ресурсов нефти на газосодержание (9–10). Каждый из коэффициентов является независимой переменной, которую можно представить в виде непрерывной функции распределения.

$$m_{geo} = BV \cdot NTG \cdot \phi \cdot S_o \cdot FVF \cdot \rho_o \quad (7)$$

$$m_{rec} = m_{geo} \cdot RF \quad (8)$$

$$VG_{geo} = m_{geo} \cdot R_S \quad (9)$$

$$VG_{rec} = m_{rec} \cdot R_S \quad (10)$$

Общий объём залежи. Различают два основных типа залежей: массивные и пластовые. Массивной называется та залежь, в которой непроницаемой является только покрывка. Внутренние пропластки внутри такой залежи гидродинамически связаны между собой.

Общий объём залежи BV зависит от геометрии пласта, его толщины H и положения контакта углеводорода с водой WC . Кровля пласта-коллектора структуры X была определена по данным геофизических исследований. Для простоты расчёта

неопределённость, связанная с точностью картирования кровли пласта, не учитывается.

Учитывая степень заполнения ловушки месторождений-аналогов, предполагается наиболее вероятным заполнение залежи на половину. Учитывая замыкающую изолинию пласта (spill point), положение контакта описывается усечённым нормальным распределением (11). Толщина пласта была определена по данным гидродинамических исследований на близлежащих месторождениях и также описывается усечённым нормальным распределением (12).

$$WC = Trunc. Norm. Inv \quad (10)$$

$$(\mu_{WC}, \sigma_{WC}^2, WC_{min}, WC_{max})$$

$$H = Trunc. Norm. Inv \quad (11)$$

$$(\mu_H, \sigma_H^2, H_{min}, H_{max})$$

Было проведено 100000 симуляций определения общего объёма залежи, в результате была получена гистограмма, представленная на рис. 2. Среднее значение полученной величины равно $\mu_{BV} = 635,6$ млн м³, стандартное отклонение $\sigma_{BV} = 400,0$ млн м³.

Полученный в результате симуляции общий объём залежи BV был аппроксимирован различными функциями распределения (рис. 3). Суммарная ошибка наименьшая у гамма-распределения. Следовательно, для моделирования функции распределения общего объёма залежи пластового типа с двумя параметрами рекомендуется использовать гамма-функцию распределения как дающую меньшую погрешность.

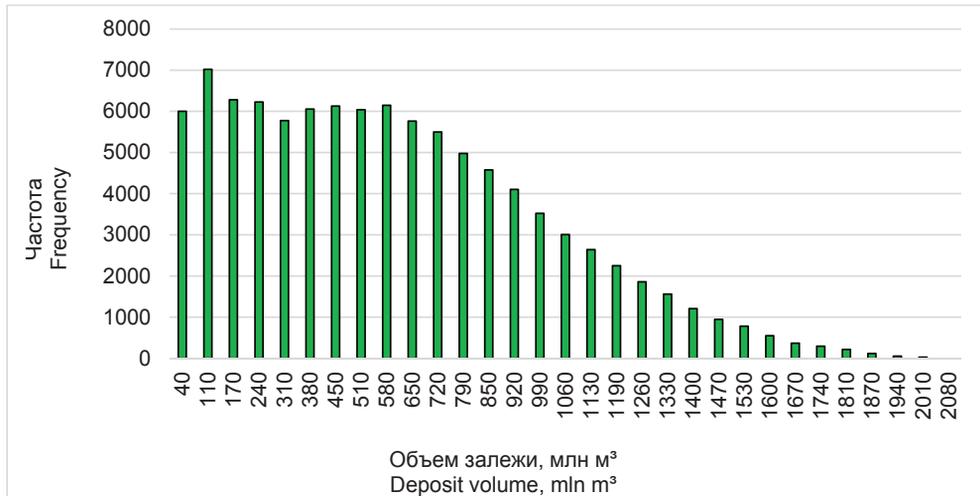


Рисунок 2. Гистограмма распределения общего объема залежи BV
Figure 2. Distribution histogram of the BV deposit total volume

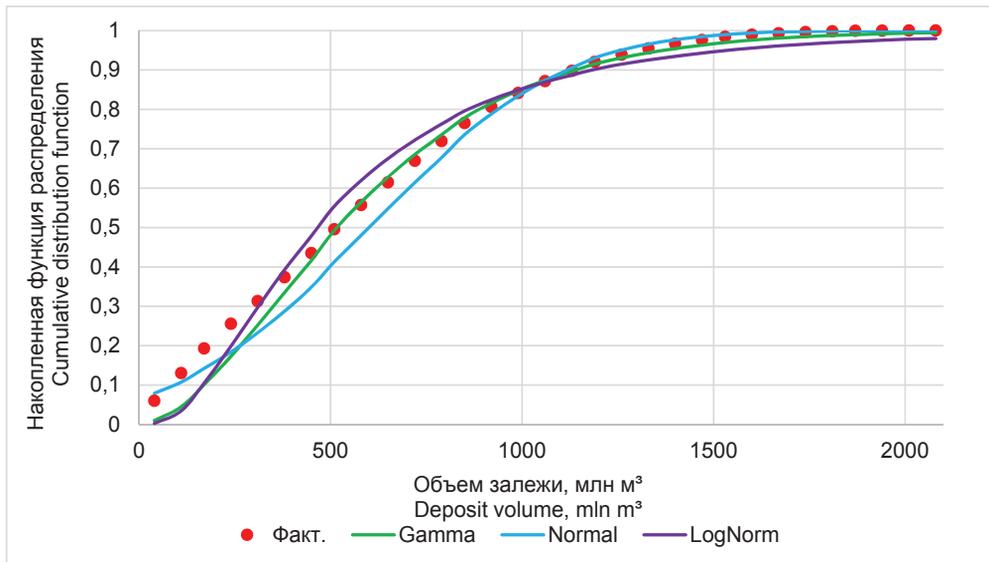


Рисунок 3. Аппроксимация общего объема залежи BV различными функциями распределений
Figure 3. Approximation of BV total bulk volume distribution by various distribution functions

Аналитический метод вероятностного подсчета ресурсов

При подсчете ресурсов объемным методом производится перемножение переменных между собой. Для аналитического расчета используется произведение двух независимых случайных величин. Пусть имеются две случайные величины X и Y со средним значением μ_X и μ_Y и дисперсиями $Var(X)$ и $Var(Y)$. Тогда произведение этих величин будет являться также случайной

величиной со средним значением и дисперсией показанными в уравнениях (1) и (12).

$$\mu_{X \cdot Y} = \mu_X \cdot \mu_Y \quad (12)$$

Далее последовательно перемножается каждая следующая переменная на результат перемножения предыдущих переменных. Вначале рассчитывается эффективный объем залежи (13, 14). После рассчитывается поровый (15, 16) и нефте-

насыщенный объёмы залежи (17, 18). Полученный нефтенасыщенный объём залежи перемножается на коэффициент усадки нефти, что даёт геологические ресурсы нефти в объёмных единицах (19, 20). Перемножив величину на плотность нефти, получают геологические ресурсы нефти в массовых единицах (21, 22). Извлекаемые ресурсы равны произведению геологических ресурсов нефти и коэффициента извлечения нефти (23, 24). И, наконец, геологические и извлекаемые ресурсы попутного газа равны произведению соответствующих ресурсов нефти на газосодержание (25, 26).

$$\mu_{NV} = \mu_{BV} \cdot \mu_{NTG} \quad (13)$$

$$\sigma_{NV}^2 = \sigma_{BV}^2 \cdot \sigma_{NTG}^2 + \mu_{BV}^2 \cdot \sigma_{NTG}^2 + \mu_{NTG}^2 \cdot \sigma_{BV}^2 \quad (14)$$

$$\mu_{PV} = \mu_{NV} \cdot \mu_{por} \quad (15)$$

$$\sigma_{PV}^2 = \sigma_{NV}^2 \cdot \sigma_{por}^2 + \mu_{NV}^2 \cdot \sigma_{por}^2 + \mu_{por}^2 \cdot \sigma_{NV}^2 \quad (16)$$

$$\mu_{HCPV} = \mu_{PV} \cdot \mu_{So} \quad (17)$$

$$\sigma_{HCPV}^2 = \sigma_{PV}^2 \cdot \sigma_{So}^2 + \mu_{PV}^2 \cdot \sigma_{So}^2 + \mu_{So}^2 \cdot \sigma_{PV}^2 \quad (18)$$

$$\mu_{Qgeo} = \mu_{HCPV} \cdot \mu_{FVF} \quad (19)$$

$$\sigma_{Qgeo}^2 = \sigma_{HCPV}^2 \cdot \sigma_{FVF}^2 + \mu_{HCPV}^2 \cdot \sigma_{FVF}^2 + \mu_{FVF}^2 \cdot \sigma_{HCPV}^2 \quad (20)$$

$$\mu_{mgeo} = \mu_{Qgeo} \cdot \mu_{\rho o} \quad (21)$$

$$\sigma_{mgeo}^2 = \sigma_{Qgeo}^2 \cdot \sigma_{\rho o}^2 + \mu_{Qgeo}^2 \cdot \sigma_{\rho o}^2 + \mu_{\rho o}^2 \cdot \sigma_{Qgeo}^2 \quad (22)$$

$$\mu_{VGgeo} = \mu_{mgeo} \cdot \mu_{Rs} \quad (23)$$

$$\sigma_{VGgeo}^2 = \sigma_{mgeo}^2 \cdot \sigma_{Rs}^2 + \mu_{mgeo}^2 \cdot \sigma_{Rs}^2 + \mu_{Rs}^2 \cdot \sigma_{mgeo}^2 \quad (24)$$

$$\mu_{VGrec} = \mu_{mrec} \cdot \mu_{Rs} \quad (25)$$

$$\sigma_{VGrec}^2 = \sigma_{mrec}^2 \cdot \sigma_{Rs}^2 + \mu_{mrec}^2 \cdot \sigma_{Rs}^2 + \mu_{Rs}^2 \cdot \sigma_{mrec}^2 \quad (26)$$

Метод Монте-Карло вероятностно-го подсчёта ресурсов

Для симуляции вероятностного подсчёта ресурсов методом Монте-Карло генерируется двумерный массив случайных чисел размером $\{m, n\}$, где m – число переменных, а n – число симуляций. Затем, используя случайное число массива в качестве значения накопленной функции вероятностного распределения CDF , по-

лучается уравнение (27), решив которое, можно получить переменную. Таким образом, решается обратная задача – нахождение аргумента, зная значение функции.

Например, параметр общего объёма залежи BV зависит от положения контакта и толщины пласта, которые в свою очередь являются переменными, моделируемыми с помощью нормального распределения. Квантиль, или значение переменной с накопленной вероятностью p , может быть выражен через обратную функцию ошибки $\text{erf}^{-1}(p)$. Обратная функция ошибки может быть рассчитана путём разложения в ряд.

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \text{Rand}(0,1) \quad (27)$$

где f_X – функция плотности распределения случайной величины X .

$$X(p) = \mu + \sigma\sqrt{2} \cdot \text{erf}^{-1}(2p - 1) \quad (28)$$

Нахождение квантилей (значения переменной) гамма-распределения производится методом подбора: принимаются две точки l и r , значение накопленной функции в которых для l меньше p , а для r больше p ($CDF(k, \theta, l) < mp$; $CDF(k, \theta, r) < p$). Затем путём итеративного приближения друг к другу находится такое значение x , для которого выполняется условие $|CDF(k, \theta, x) - p| \leq \text{eps}$, где eps – заданная точность переменной.

Проведя анализ над полученной выборкой, используя формулы из табл. 1, можно получить данные о статистическом распределении функции. Перцентили P10, P50, P90 рассчитываются либо по уравнению (3), используя вероятности 0,9, 0,5 и 0,1, либо путём ранжирования и разбиения множества полученных результатов.

Результаты

Случайная выборка или латинский гиперкуб

Для сравнения методов выборки случайных чисел была выбрана простая функция произведения двух случайных величин вида $Y = X_1 \cdot X_2$. Было проведено 20 расчётов с 10, 100, 1000 и 10000 симуляций, с использованием случайной выборки и с помощью латинского гиперкуба. Затем были рассчитаны ошибки среднего и среднеквадратичного отклонения относительно результатов, рассчитанных аналитическим способом. Результаты расчёта представ-

лены на рис. 4. Как видно из рисунка, латинский гиперкуб во всех вариантах даёт меньшую погрешность, по сравнению со случайной выборкой (около 2 раз). Однако необходимо учитывать, что данный результат связан с несовершенством генератора случайных чисел. В настоящее время большинство генераторов случайных чисел завязаны на алгоритме псевдослучайных чисел, которые, в свою очередь, зависят от текущего такта центрального процессора компьютера.

Аналитическое решение

Формулы для аналитического решения представлены в уравнениях (13–26).

Однако данные уравнения были модифицированы, т.к. в качестве входных данных использовались усечённые распределения. В табл. 4 представлены результаты аналитических расчётов. Полученные результаты использовались для сравнения с результатами, полученными путём симуляции Монте-Карло при разном количестве симуляций.

Таблица 4. Результаты аналитического вероятностного подсчёта ресурсов
Table 4. Results of analytical probabilistic resources estimation

Параметр Parameter	Среднее Mean	Средне- квадратичное отклонение Standard deviation	MIN	MAX	Среднее усе- чённое Truncated mean	Среднеква- дратичное отклонение Truncated standard deviation
Общий объём пород, тыс. м ³ Total rock volume, thousand m ³					635601737	400048056
Доля коллектора, д. ед. Net to Gross ratio, units	0,70	0,08	0,60	0,80	0,70	0,05
Коэффициент пористости, д. ед. Porosity, units	0,20	0,03	0,15	0,25	0,20	0,02
Коэффициент нефтенасыщенности, д. ед. Oil saturation coefficient, units	0,68	0,10	0,60	0,75	0,68	0,04
Пересчетный коэффициент, д. ед. Formation volume factor, units	0,80	0,05	0,70	0,90	0,80	0,04
Плотность нефти, т/м ³ Oil density, t/m ³	0,85	0,05	0,70	0,90	0,84	0,04
Геологические ресурсы нефти, тыс. т Geological resources of oil, thousand tons					40186295	26558156
Коэффициент извлечения нефти, д. ед. Oil recovery factor, units	0,30	0,05	0,20	0,40	0,30	0,04
Извлекаемые ресурсы нефти, тыс. т Recoverable oil resources, thousand tons					12055888	8244297

Число симуляций

Были проведены серии расчётов аналитическим методом и методом имитационного моделирования для различного числа симуляций. Для исключения влияния генератора случайных чисел расчёт повторялся 50 раз, каждый раз генерировались новые массивы случайных чисел. Анализировалось среднее значение погрешности при 10, 50, 100, 200, 500, 1000, 10000, 50000 и 100000 симуляций подсчёта ресурсов (табл. 5).

Оценивалась погрешность результатов относительно аналитического решения, которое даёт точный ответ. Увеличение числа симуляций повышает точность вычислений. Кроме того, расчёты чувствительны к генератору случайных чисел, из-за чего возможны случаи, когда большое число симуляций может снизить точность. Эмпирическим путём были получены формулы, с помощью которых можно определить требуемое число симуляций (рис. 5).

Таблица 5. Влияние количества симуляций на ошибку при симуляции Монте-Карло
Table 5. Effect of the number of simulations on the Monte Carlo simulation error

Параметр Parameter		Аналитическое решение Analytical Solution	Кол-во симуляций Монте-Карло Number of Monte Carlo simulations					
			10	100	1000	10000	50000	100000
1 реализация 1 simulation	Среднее значение извлекаемых ресурсов нефти, тыс. т Mean of recoverable oil resources, thousand tons	12055888	8222317	12273603	11924966	12016274	12071106	12048959
	Среднеквадратичное отклонение извлечённых ресурсов нефти Standard deviation of recoverable oil resources	8244297	5547411	7475631	8107110	8352298	8251386	8190219
	Погрешность данной реализации Error of this implementation	0,0	0,3180	0,0181	0,0109	0,0033	0,0013	0,0006
ИТО- ГИ RE- SULTS	Среднее значение ошибки Average error		0,1582	0,0562	0,0201	0,0055	0,0023	0,0018
	Среднеквадратичное отклонение ошибки Standard deviation of error		0,1363	0,0420	0,0144	0,0044	0,0017	0,0013
	V2		0,7426	0,5582	0,5135	0,6206	0,5843	0,5364

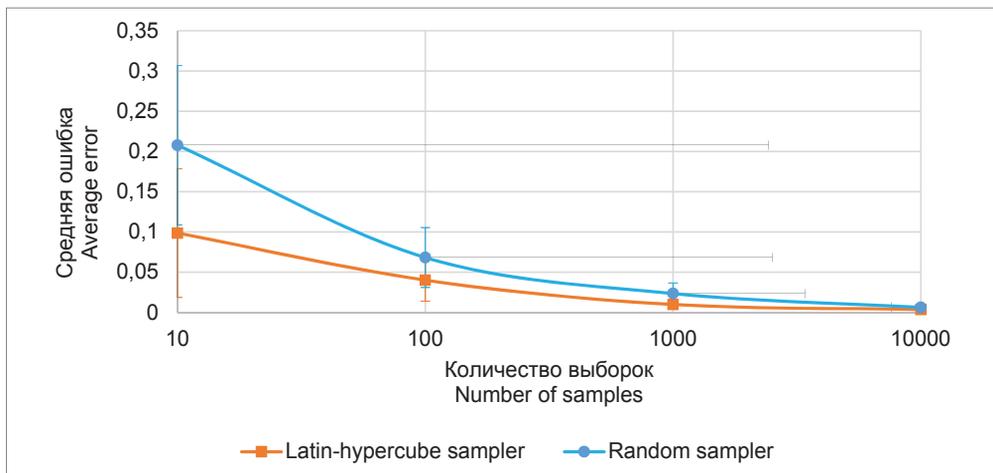


Рисунок 4. Сравнение методов выборки случайного числа
Figure 4. Comparison of random number sampling methods

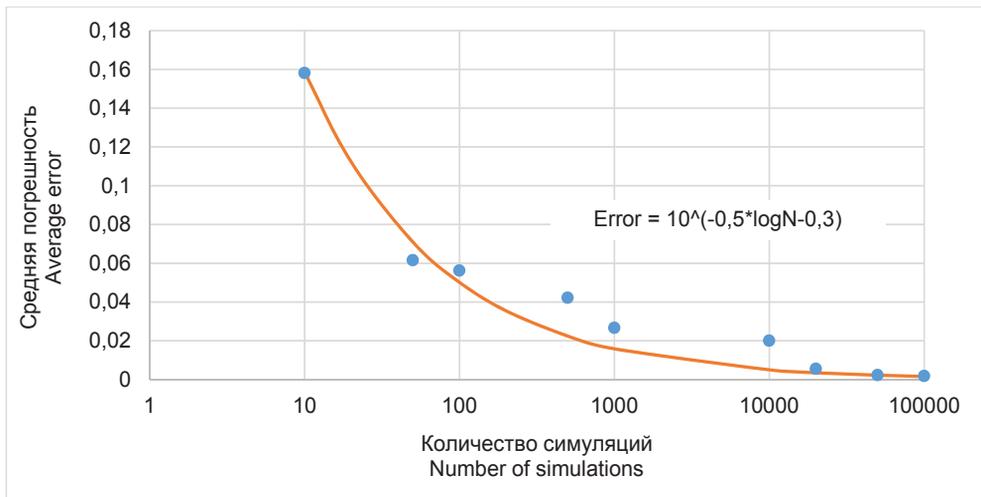


Рисунок 5. Средняя погрешность при разном количестве симуляций
 Figure 5. Average error for different number of simulations

Выводы

Аналитический расчёт при вероятностном подсчёте ресурсов позволяет с наименьшими вычислительными затратами найти точную функцию распределения геологических и извлекаемых ресурсов. Однако при использовании аналитического расчёта возникают трудности с выводом формул, а также высокой вероятностью допущения технической ошибки при проведении вычислений.

Симуляция Монте-Карло позволяет приближенно находить решение для сложных функций, в которых может быть большое количество переменных. Точность решения зависит от количества симуляций, однако после определенного порога дополнительное увеличение количества симуляций практически не увеличивает точность расчёта.

Имеющиеся в настоящее время генераторы случайных чисел, основанные на алгоритме псевдослучайных чисел, не являются совершенными. Латинский гиперкуб в качестве метода выбора при определении случайных величин позволяет в

некоторой степени уменьшить влияние генератора случайных чисел. По результатам расчётов, латинский гиперкуб показывал точность большую в среднем в 2 раза, по сравнению с использованием метода случайной выборки. С другой стороны, латинский гиперкуб, в свою очередь, требует дополнительных вычислений, что может быть скомпенсировано большим количеством симуляций с использованием метода случайной выборки.

По результатам работы были выведены формулы, позволяющие определить количество симуляций для получения функций распределения геологических или извлекаемых ресурсов с заданной точностью. Например, если необходима точность порядка 1%, то необходимо проведение порядка 10000 симуляций. В большинстве случаев такой точности, учитывая неопределенность входных данных, будет достаточно. Важным является исключение значимого влияния метода расчёта на итоговый результат, т.е. обеспечение повторяемости результатов.

ДОПОЛНИТЕЛЬНО

Источник финансирования. Автор заявляет об отсутствии внешнего финансирования при проведении исследования.

Конфликт интересов. Автор декларирует отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

ADDITIONAL INFORMATION

Funding source. This study was not supported by any external sources of funding.

Competing interests. Author declare that he has no competing interests.

REFERENCES

1. Metropolis N, Ulam S. The Monte Carlo Method. *Journal of the American Statistical Association*. 1949;44(247):335–341.
2. Shiryayev AN. Probability. Moscow: Nauka; 1980.
3. Iman RL. Latin Hypercube Sampling. *Wiley StatsRef: Statistics Reference Online*. 2014. doi:10.1002/9781118445112.stat03803.
4. Dekking FM, Kraaikamp C, Lopuhaa HP, Meester LE. *A modern introduction to probability and statistics: understanding why and how*. London: Springer; 2005.
5. Christian W (Particle Physics Group Fysikum University of Stockholm). Hand-book on Statistical Distributions for experimentalists. *Internal Report*. Stockholm; 1996. Report No.: SUF-PFY/96-01.

ИНФОРМАЦИЯ ОБ АВТОРЕ

Садыков Раман Муханович
ORCID 0000-0002-5936-2036
e-mail: r.sadykov@niikmg.kz.

AUTHOR'S INFO

Raman M. Sadykov
ORCID 0000-0002-5936-2036
e-mail: r.sadykov@niikmg.kz.