**1.** Постановка задачи

* 1. *Геофизическая постановка задачи*

Газовая скважина исследована аппаратурой мультиметодного многозондового нейтронного каротажа (ММНК) и по методике ММНК-Кг определены *N* значений коэффициента газонасыщенности *Кг* в разных пластах разреза. После этого с целью проверки корректности всей технологии проводятся испытания скважины на приток, в которых на качественном уровне измеряется состав фактически добываемой продукции газоводяной смеси. Это означает, что продукция классифицируется на небольшое число градаций M ~ 3 - 5 разбиений шкалы *Кг* на эквидистантные широкие интервалы протяженностью по Δ*Кг*= [*Кг*] / *M*, где [*Кг*] – максимально возможный диапазон изменения *Кг* в исследуемых геолого-промысловых условиях. Например, наиболее часто используемыми интервалами изменения Δ*Кг* для типовых [*Кг*] = [0,1], *M* = 4 являются по терминологии газовиков: «вода (0-0.25), вода+газ (0.25-0.5), газ+вода (0.5-0.75), газ (0.75-1)». Затем проверяются доли *p*0, % правильных попаданий предсказанных ММНК-Кг численных значений *Кг* в соответствующие им широкие интервалы Δ*Кг*. Если большинство этих долей *p*0 > 80-90 %, то технология признается корректной, т.к. получила качественное подтверждение по результатам испытаний, считающихся одним из наиболее прямых и убедительных способов тестирования методик в скважинной геофизике.

Представляется очевидным, что при достаточно большом числе определений *Кг* и, разумеется, при условии выполнения достаточно жесткого критерия подтверждаемости по *p*0 любая разумно введенная оценка фактической средней погрешности определения *Кг* должна дать величину, существенно меньшую широких интервалов разбиения Δ*Кг*. Другими словами, это означает, что технологию ММНК можно будет переквалифицировать из качественной по способу ее подтверждения в количественную по фактически достигаемому уровню погрешности определения *Кг*.

Поэтому задачами работы явились следующие:

* численное обоснование этого утверждения на основе теории вероятностей с разработкой алгоритма и программы расчета фактической средней погрешности определения \emph{Кг} по всем имеющимся данным определений и подтверждений в исследуемой скважине;
* численное изучение поведения погрешности в зависимости от варьируемых параметров $M$, $N$, [\emph{Кг}], $p\_0$ и характера вероятностного распределения найденных значений \emph{Кг} на [\emph{Кг}], $P$(\emph{Кг}) - от равномерного до гауссового с большой дисперсией. Диапазоны изменения варьируемых параметров:
  + *M = 3, 4, 5*
  + *[\emph{Кг}] ~ [0.5, 1], [0.25, 1], [0, 1]*
  + *p*0 = 80%; 90%;
* выдача практических рекомендаций по выбору единственного управляемого параметра $N$ в зависимости от априори задаваемых геофизиками и газовиками параметров $M$ и [\emph{Кг}], а также от фактически получившихся характеристик – распределения $P$(\emph{Кг}) и значений $p\_0$ в результате сопоставления определений и подтверждений \emph{Кг}.

*1.2 Математическая формулировка*

Цель -- предсказать погрешность $\varepsilon\_0$ выдаваемых нашей программой значений $x$ искомого параметра $y$.

Область изменения $x \in $ [\emph{Кг}] разбита на интервалы, для каждого из которых есть *N*j экспериментов. Считается, что искомая погрешность $\varepsilon\_0$ может меняться от интервала к интервалу, но постоянная внутри интервала (т.е. точнее писать $\varepsilon\_{0j}$). Фиксируем $j$ и работаем в выбранном интервале, поэтому далее индекс интервала $j$ опущен.

Есть $N$ экспериментов, про которые известно, что в каждом из них истинное значение $y\_i$ попало в интервал, т.е.   . На каждый из этих экспериментов у нас есть результат работы нашей программы $x\_i$. Предполагается, что истинное значение $y\_{i}$ распределено по Гауссу со средним $x\_i$ и некой дисперсией $\varepsilon$, т.е.



где  - вероятность того, что *A* верно, а



- Гауссово распределение.

Ищем зависимость



такую, что вероятность реализации описанной выше ситуации (т.е. что все истинные значения попали в интервал) $ = p\_0$, т.е.



Из сторонних соображений считается известным минимально возможная погрешность $\varepsilon\_{min}$, т.е. если метод выдает $\varepsilon\_0 < \varepsilon\_{min}$, то считаем $\varepsilon\_0 = \varepsilon\_{min}$.

2. Предлагаемое решение

*2.1 Идея и приближения*

Задав $\varepsilon$, можно посчитать вероятность реализации ситуации, описанной в постановке -- попадания всех истинных значений параметра $y\_i$, распределенных по Гауссу каждый около своего $x\_i$, в интервал $[a;b]$. Далее предположение - эта вероятность равна нашей целевой вероятность $p\_0$. Не очевидно, почему это должно выполняться точно (скорее всего это не выполняется), но для оценки предложено использовать такую модель.

Поясним разумность данного выбора. Будем брать пробные $\varepsilon$ и смотреть как от этого зависит ожидаемое поведение истинных значений $y\_i$ относительно наших точек $x\_i$. Для примера возьмем весь интервал $[0.2; 0.5]$ и предположим что у нас имеются 5 точек, для которых наша программа выдала ответы 0.22, 0.3, 0.35, 0.4, 0.43. Если предположить, что погрешность наший предсказаний $\varepsilon = 0.02$, то плотность вероятности для каждого из 5 истинных значений будет выглядет так:

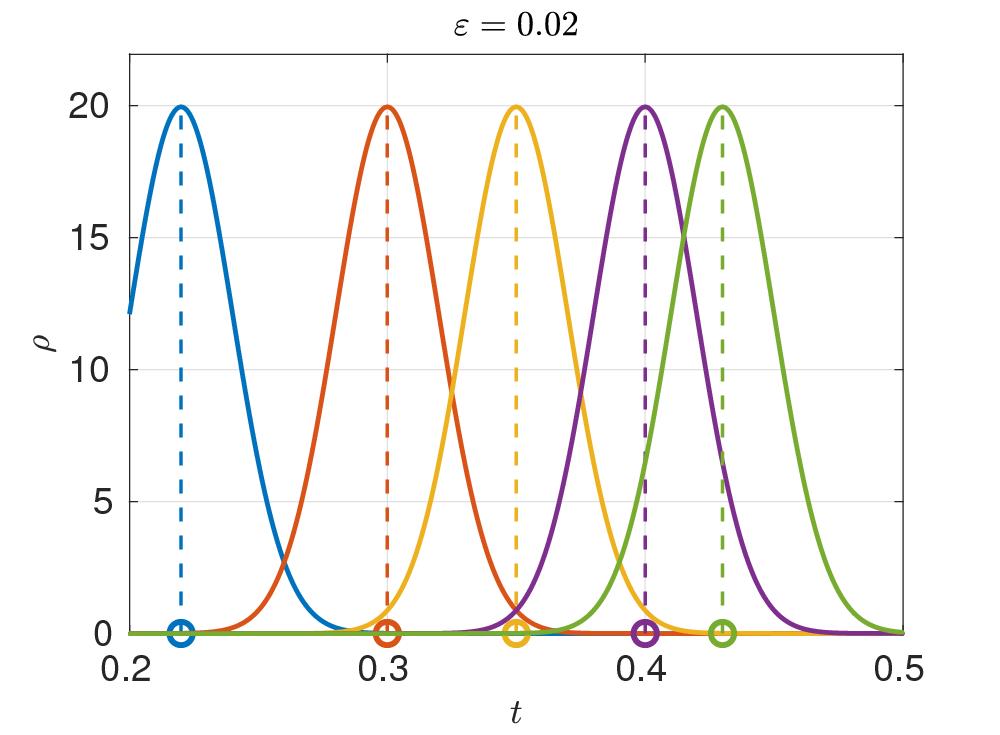


Рис. 1: Разным цветам отвечают разные эксперименты. Для каждого эксперимента: проколотый круг на оси Х - наше предсказание ответа, купол - распределение плотности вероятности того, что истинное значение ответа примет значение *t*, в зависимости от *t*.

На Рис. 1 видно, что для всех точек кроме синей почти вся кривая (вероятность = площадь под кривой) находится в исследуемом интервале. Это значит, что при $\varepsilon = 0.02$ для всех точек кроме синей вероятность того, что интинное значение параметра попадет в интервал, равна почти 100\%. Синяя же точка находится на расстоянии $\sim 1 \sigma$ (в данном случае 0.02), что значит, что вероятность того, что истинное значение параметра в синем эксперименте попадет в интервал $[0.2; 0.5]$ будет $\approx 16 \%$. Попадания истинных значений в интервал -- события независимые, поэтому вероятность реализации картины в целом будет произведением вероятностей попадания каждого значения в интервал по отдельности. В нашем случае все вероятности кроме синей $\approx 1$, поэтому общая вероятность $\approx$ синяя вероятность $\approx$ 84\%. Это значит, что если бы погрешность нашей программы была 0.02, то вероятность случатся тому что случилось на рассматриваемых 5 экспериментах в совокупности была бы 84 \%. Поняв это, можно решить обратную задачу: сказать, что мы верим эксперименту на скажем 95\%, и найти такое $\varepsilon$, при котором вероятность его реализации как раз будет 95\%. Понятно, что задача решаема -- если в предыдущем примере мы возьмем $\varepsilon = 0.008$, то распределение вероятностей для истинных значений будет

