



1 Апреля 2012 г.

Презеденту Micro-Tracers, Inc.

А. Айзенбергу,

1370 Ван Дайк Авеню

Сан-Франциско, Калифорния 94124

Г-н Айзенберг,

Настоящее письмо является ответом на Ваш запрос об обзоре и оценке общих статистических методов проверки уровня однородности смеси кормов на основе оценок распределения микротрейсерных частиц. Я просмотрел предоставленную вами литературу, в частности вашу статью 1976 года “Использование микротрейсеров для определения однородности. Формула кормов для животных” и отчет 2010 года, подготовленный компанией TNO Science and Industry. Я прилагаю копию моего полного отчета, который содержит более подробные сведения, чем краткое изложение, представленное здесь.

Процедура, описанная в вашей статье, и теория вероятностей, объясненная в ней, точны и подходят для проверки однородности смеси. Использование как пуассоновских, так и хи-квадратных тестов теоретически обосновано, исходя из допущений и условий смешивания кормов. Я думаю, что из вашего письма и из наших бесед ясно, что вы твердо понимаете, как каждый аспект процесса смешивания может повлиять на достоверность и точность результатов теста.

С моей точки зрения, следующие вопросы представляют особый интерес, когда речь заходит об использовании частиц микротрейсера в процессе смешивания кормов:

1. Сколько образцов должно быть взято из смеси для анализа?
2. Сколько частиц должно быть добавлено в смесь изначально, чтобы обеспечить точные результаты испытаний?
3. Насколько большим должен быть каждый образец (измеряется в граммах)?
4. Откуда в смеси должны быть взяты образцы?
5. Сколько суб-образцов должно быть взято из каждого образца?

Тема, которая всплыла в моем исследовании этих вопросов, заключается в том, что нет единого ответа ни на один из этих вопросов. Как и в большинстве статистических тестов и выводов, количество выборок очень важно, так как оно приводит к сходимости результатов и указывает,

подходят ли наши тесты для интерпретации данных. Из этого следует, что, поскольку увеличение количества образцов предпочтительнее над их уменьшением, ответ на большинство этих вопросов должен быть “как можно больше”. Тем не менее, я думаю, что крайне важно, чтобы мы сначала рассмотрели ответ на более фундаментальный вопрос: насколько полным должно быть смешивание?

Это правда, что тестирование большего количества образцов или добавление большего количества частиц может только улучшить наши результаты, но это улучшение может быть неуместным во многих случаях. Я полагаю, исходя из контекста, в котором проводится это тестирование, что статистическая теория не должна использоваться для формулирования политики, касающейся предельно допустимых размеров выборки и т.д. Вместо этого теория вероятностей лучше подходит для установления верхней и нижней границ параметров нашей процедуры. Пока мы находимся в этих пределах, возможно, лучше всего позволить экономическим факторам—таким как стоимость смешивания, стоимость ингредиентов или затраты на производство неполной смеси—определить, как эти параметры лучше настроить. Именно по этой причине я предлагаю более интуитивный подход к процедурам тестирования по сравнению с аналитически сложными методами, предложенными в отчете TNO.

Имея это в виду, я рекомендую протестировать не менее 10 образцов примерно по 100 г каждый, содержащих от 50 до 100 частиц, которые были взяты из нескольких мест внутри смесителя в нескольких различных периодах смешивания. Большая точность может быть получена путем увеличения этих чисел; более высокая эффективность (более низкая стоимость тестирования) достигается за счет их снижения. Достижение компромисса между этими вариантами в конечном счете является самой большой проблемой, которую необходимо решить. Использование суб-образцов не сильно улучшит наши измерения и должно использоваться только в том случае, если стоимость этого близка к нулю.

Я не думаю, что какие-либо из этих рекомендаций отличаются от того, что вы сейчас считаете правильным.

Мне было очень приятно работать с вами над этим проектом. Ваше знание производственных процессов и теории, на которых они основаны очень впечатляет, и я нахожу этот предмет очень интеллектуально стимулирующим. Пожалуйста, не стесняйтесь обращаться ко мне в любое время, если я могу быть вам чем-то полезен или если что-то здесь или в моем отчете неясно.

Искренне Ваш,

Handwritten signature of David Bant in black ink, written in a cursive style.

Кандидат наук Дэвид Бернотас

Об использовании распределений твердых частиц для определения степени однородности кормовой смеси

Дэвид Бернотас, кандидат наук.
Кафедра экономики
Калифорнийский университет, Сан-Диего
1 апреля 2012 года

Microtracers Inc. обратилась ко мне с просьбой обобщить соответствующие статистические методы определения качества смешивания партии корма на основе расчетного распределения трейсерных частиц¹.

Основным документом, на котором данная работа основана является статья Дэвида Айзенберга из Microtracers Inc. датированная 1976 годом под названием «Использование микротрейсеров для определения однородности. Формула кормов для животных », где предлагается (правильно), что распределение Пуассона в сочетании с критерием Хи-квадрат Пирсона может использоваться для оценки полноты (однородности) смешивания. Хотя статистическая теория в основе проблемы не возникает сомнений, некоторые вопросы действительно возникают в контексте извлечения микротрейсеров из смеси кормов для определения степени ее однородности.

Например:

1. Сколько проб нужно отобрать из смеси для анализа?
2. Сколько частиц следует добавить в смесь изначально, чтобы гарантировать точные результаты?
3. Какого размера должен быть каждый образец (в граммах)?
4. Откуда следует брать пробы?
5. Сколько подвыборок следует отобрать из каждой пробы?

Я предложу ответ на каждый из них. Я говорю "предложу ответ", потому что есть несколько эффективных способов решения этих проблем - каждый имеет свои издержки и преимущества - и лучшая альтернатива будет зависеть от ситуации и предпочтений конкретного человека. Я сделаю все возможное, чтобы выделить компромисс между стоимостью и выгодой для каждого варианта.

1 Основы теории вероятности

Этот раздел и его подразделы предназначены для полноты изложения и обеспечения единообразия обозначений и терминологии.

¹ Терминология здесь является деликатным делом. Смесь является "полной", если она соответствует стандартам и предпочтениям человека, смешивающего корм, исходя из контекста. "Хорошее" и "плохое" определены менее точно, но я буду использовать "хорошее" и "полное" взаимозаменяемо, так же как "плохое" и "неполное". "Однородный" означает, что смесь настолько случайна, насколько это возможно — технически это называется "максимальной энтропией". Таким образом, обозначения степени смешивания корма от наиболее смешанного к наименее смешанному выглядят следующим образом:
"однородное" > "полное" = "хорошее" > "неполное" = "плохое"

1.1 Распределения Пуассона

Распределение Пуассона подходит для моделирования распределения редких, дискретных событий в течение определенного периода времени, расстояния или объема.

Классические примеры:

количество телефонных звонков в час (время), количество мутаций на нити ДНК (расстояние) и скопление галактик во Вселенной (объем). (См. Примеры в Hammersly (1972), Saslaw (1989) и Норе (1979) для примера и строго подхода)

Распределение частиц в однородном кормовой смеси будет соответствовать распределению Пуассона.

Напомним свойства распределения Пуассона:

$$\text{prob}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (1)$$

где k - число наблюдаемых частиц, а λ - ожидаемое число частиц (среднее) и дисперсия. Отсюда следует, что коэффициент вариации (CV), который связывает дисперсию со средним значением распределения определяется следующим образом:

$$CV = \frac{\sqrt{\lambda}}{\lambda} = \frac{1}{\sqrt{\lambda}} \approx \frac{1}{\sqrt{X}}. \quad (2)$$

Асимметрия и эксцесс (пикообразность) распределения Пуассона, так же полностью определяются параметром λ :

$$\text{Асимметрия} = \frac{1}{\sqrt{\lambda}} \quad (3)$$

$$\text{Эксцесс} = 3 + \frac{1}{\lambda}$$

Поскольку $\lambda \geq 0$ значения асимметрии всегда положительны (распределение наклоняется влево от центра) и значения эксцесса всегда ≥ 3 (распределение выше, чем нормальное распределение). Поскольку среднее значение, дисперсия, асимметрия и пик Пуассона являются характеризуется значением λ этот параметр сообщает нам все, что нам нужно знать о форме и расположении распределения Пуассона. Если λ достаточно велико ($\lambda \approx 1000$), то форму распределения Пуассона можно очень точно аппроксимировать, используя нормальное распределение. Для значений значительно меньших ($\lambda \approx 10$) распределение Пуассона все еще можно аппроксимировать нормальным распределением, если корректировка производится при вычислении среднего и дисперсии. Эта корректировка называется «поправкой Йетса», но на самом деле это не более чем обман, поскольку, вы знаете, что маленький λ приводит к перекосу влево, то корректируете наклон вправо. Здесь нам это исправление не понадобится.

Примечание о сходимости: на протяжении всей статьи я буду делать такие утверждения, как «если λ достаточно велика» или «это верно, если $\lambda > 10$ » и т.д. Эти утверждения сделаны на основе признанных результатов о сходимости распределений. Например, Закон больших чисел гласит, что

выборочное среднее будет стремиться к истине если взято достаточное количество образцов. Нет четкого ограничения для "достаточного количества образцов". Сходимость описывает только направление движения - фактически же мы никогда не достигнем цели. Больше количество образцов или большая λ смогут улучшить результаты, но это полностью зависит от личных предпочтений и контекста, чтобы определить, когда достаточно. Это сложно в такой обстановке, потому что принятие решения о таких вещах, как размер выборки и количество микротрейсеров являются частью производственного процесса. Я не могу ответить на эти вопросы или утверждать, что правильно, - я могу комментировать только то, что разумно. По сути, мы знаем только то, что "ужасно". Все остальное просто «не ужасно».

1.2 Тестирование

В конечном счете, мы будем использовать приведенную выше теоретическую модель для сравнения, чтобы определить соответствие наблюдаемых нами образцов распределению Пуассона.

Визуально мы можем представить, как точка за точкой строим график, используя данные наших измерений, поверх теоретического графика распределения Пуассона и видим, как все выстраивается.

(На самом деле статистические таблицы говорят нам, как все "выстраивается", и это подтверждает оригинальная статья Дэвида Айзенберга и следующий раздел.) Если они действительно выстраиваются так как должно быть—они пуассоновские, если нет, то требуется дополнительное смешивание. Ключевым моментом является уверенность в том, что частицы неизбежно будут пуассоновскими в смеси, если перемешивание будет достаточно тщательным. Это закон природы.

Подумайте об этом так: в примере с кормовой смесью есть два распределения. Первое - это истинное распределение Пуассона, которому в конечном итоге будут соответствовать частицы трейсера после смешивания. Форма этого истинного распределения определяется истинным λ , которое является функцией количества добавленных частиц микротрейсера и объема загрузки корма в смеситель (уравнение(4)). Второе распределение - это фактическое распределение частиц в смесителе. В определенные моменты процесса смешивания (особенно на ранней стадии) я понятия не имею, как выглядит это фактическое распределение. Чтобы оценить его, я беру образцы. В последующих тестах будут использоваться образцы, чтобы определить, когда фактическое распределение достаточно близко к истинному. Другими словами, эти методы проверяют, находятся ли частицы, наконец, там, где они должны быть. Диаграмма в следующем разделе иллюстрирует это.

Контекст кормовой смеси здесь очень уникален для статистического моделирования, потому что мы знаем, каково фактическое значение λ . Как правило, это неизвестно. (Сколько галактик во Вселенной?) Это знание позволяет мне использовать слово "истинный" для описания теоретического распределения частиц (обычно это называется "популяционным" распределением). Вычислить истинное λ очень просто:

$$\lambda = \frac{\text{кол-во добавленных частиц}}{\text{общий объем корма}} \quad (4)$$

Конечно, есть некоторая неопределенность в обоих этих значениях. Для микротрейсеров мы знаем объем в котором они добавлены в корм (или добавленную массу), и мы можем использовать это для вычисления числа частиц, но, очевидно, существует некоторая погрешность измерения. Точно так же расчет общего объема (или массы) корма включает в себя некоторую ошибку.

Точность и аккуратность нашего расчета λ будут рассмотрены позже в этой статье. Я хочу повторить замечание, сделанное в сноске ранее в этой статье. "Однородная" смесь означает, что частицы,

наконец, достигли своего распределения Пуассона—дальнейшее смешивание не увеличит случайность. "Полная" или "хорошая" смесь означает, что частицы еще не совсем пуассоновские, но они достаточно случайны для целей, с которыми сталкивается человек, смешивающий корм. "Однородный" - настолько случайный, насколько это возможно, "полный" - настолько случайный, насколько это необходимо.

1.2.1 Критерий Пуассона для определения формы наблюдаемого распределения

Напомним, что форма распределения Пуассона полностью описывается параметром λ . Представленный здесь критерий Пуассона предназначен для оценки значения этого параметра для наблюдаемого распределения, и при этом он будет определять "форму" наблюдаемого распределения. Поскольку λ также является средним значением, этот тест также оценивает положение наблюдаемого распределения, но я хочу сосредоточиться на форме, чтобы подчеркнуть разницу между этим тестом и тестом хи-квадрат. Итак, предположим, что тест Пуассона предназначен для формы—он отвечает на следующий вопрос: "Имеет ли график полученных данных форму Пуассона?"

Если смесь однородна, каждый образец, взятый из смесителя, должен соответствовать распределению Пуассона с известным параметром λ . Чтобы проверить это для одного образца, вычислите следующее значение, используя количество частиц микротрейсера X и известное значение для λ :

$$\frac{\lambda^x e^{-\lambda}}{X!} \quad (5)$$

Это даст мне вероятность того, что я увижу X , учитывая тот факт, что λ истинно. Например, пусть $\lambda=100$ и предположим, что я насчитал 90 частиц из образца. Уравнение (5) дает вероятность 0,025, что означает, что вероятность того, что я увижу 90 частиц при распределении Пуассона с $\lambda = 100$, составляет всего 2,5%. Это может показаться небольшим, но имейте в виду, что есть миллионы возможных исходов для каждого образца—я мог бы получить количество частиц в диапазоне между нулем и общим количеством частиц, добавленных в смеситель. На самом деле при $\lambda = 100$ наибольшая вероятность в 4% возникает только в одном случае - при $X = 100$. Чтобы сделать наш тест более надежным, мы выбираем диапазон значений и определяем вероятность того, что X попадает в этот диапазон. Например, учитывая $\lambda = 100$ каковы шансы, что $90 \leq X \leq 110$? Оказывается, что вероятность этого составляет около 70,6% (найденная путем взятия суммы индивидуальных вероятностей, найденных по уравнению (5) для $X = 90, X = 91...$ и так далее). Расширение этого диапазона увеличит вероятность. Например, если я хочу, чтобы с вероятностью 90% X находился в определенном диапазоне (учитывая $\lambda = 100$), диапазон вокруг X должен быть $84 \leq X \leq 116$.

Таким образом, построены таблицы вероятностей, подобные тем, которые использовались в оригинальной работе Айзенберга. Дальнейшее улучшение состояло бы в том, чтобы использовать несколько образцов, взять их среднее значение, и использовать это среднее значение вместо единичного наблюдения. Это очень похоже (хотя технически и не идентично) на процедуру, описанную в статье Айзенберга. Там он вычисляет среднее значение нескольких выборок, и использует это среднее значение для определения диапазона для λ (в первом случае определяется диапазон вокруг X , а втором-диапазон вокруг λ)

Любой из этих методов допустим, и какой использовать дело вкуса—хотите проверить вероятность определенной λ имея данные выборки или хотите проверить вероятность определенной выборки имея данные о λ ?

В контексте кормовых смесей уравнение (4) гласит, что я должен иметь хорошее представление о том, чему равно λ , если я знаю количество добавляемых частиц и знаю количество смешиваемого

корма. Очевидно, что неопределенность в любом из этих параметров будет превышать неопределенность в моем знании λ . Точное знание λ может привести меня к тесту, который определяет доверительный интервал вокруг значения выборки X или \bar{X} , с другой стороны, если у меня есть уверенность в моей точности подсчета значений выборок, но я не уверен в том, сколько частиц я добавил (или что случилось с ними после того, как я их добавил), тогда я могу вместо этого предпочесть проверить доверительный интервал вокруг λ , используя известное мне значение X или \bar{X} .

Здесь есть опасность: если я использую выборочное значение X (или выборочное среднее значение \bar{X}) для построения доверительного интервала я могу сказать что-то вроде этого: "Учитывая \bar{X} , истинное среднее значение "будет находится между А и В в 90% случаев". Я не могу сказать так: "учитывая \bar{X} , то 90% образцов попадут между А и В". Первый включает в себя вывод о популяции на основе выборки—стандартная статистическая практика. Последнее включает в себя вывод о будущих образцах, основанный на текущей выборке. Это недопустимо. Я могу либо предсказать истину (λ) на основе выборки (\bar{X}), либо я могу предсказать выборку, зная истину.

Я не могу предсказать выборку, основанную непосредственно на другой выборке. В статистике мы обычно вынуждены предсказывать истину на основе выборки, потому что истина обычно неизвестна. В задаче о кормовой смеси мы находимся в уникальном положении, зная, что такое λ на самом деле, так как мы добавляем известное количество частиц микротрейсеров, поэтому мы можем сделать любой прогноз. Это роскошь контекста.

Вероятность того, что наблюдаемое X попадет в интервал от \underline{X} до \bar{X} , при известном λ , такова:

$$\text{Вероятность, что } X \in [\underline{X}, \bar{X}] = \sum_{\underline{X}}^{\bar{X}} \frac{\lambda^x e^{-\lambda}}{x!} \quad (6)$$

Общая формула для доверительного интервала вокруг λ :

$$\frac{\chi^2(\alpha/2; 2k)}{2} \leq \lambda \leq \frac{\chi^2(1-\alpha/2; 2k+2)}{2} \quad (7)$$

Где $\chi^2(p; n)$ есть χ^2 для уровня значимости p (например 5% или 1%) и степеней свободы n .

1.2.2 Критерий хи-квадрат Пирсона (χ^2) для определения положения наблюдаемого распределения

Поскольку наши наблюдения, вероятно, будут независимыми и случайными, если процедура тестирования разумна, то для определения, согласуется ли наблюдаемое распределение частиц трейсера с теоретическим распределением Пуассона целесообразно использовать критерий хи-квадрат Пирсона. Это называется тестом "степени соответствия"—насколько хорошо наблюдения "соответствуют" нашей теории? Хотя распределение Пуассона не является непрерывным распределением, принято считать, что в случае, когда λ достаточно велико, не требуется корректировка из стандартного критерия хи-квадрат. Опять же, коррекция Йейтса - это вариант (который мы не будем использовать). Я прокомментирую, что означает "достаточно большой", в следующем разделе. (Уравнение (8) справедливо для $\lambda > 5$)

Критерий хи-квадрат Пирсона прост. Если предположение о независимости верно, то:

$$\sum_{i=1}^N \frac{(X_i - \lambda)^2}{\lambda} \quad (8)$$

соответствует распределению χ^2 . Процедура состоит в том, чтобы просто подсчитать частицы в каждом образце, использовать известное значение λ для вычисления суммы указанной выше и посмотреть на таблицу хи-квадрат, чтобы определить вероятность увидеть результат, который мы получили, если лежащие в основе распределения Пуассона характеристики видны в наших данных.

Низкие вероятности (малые р-значения) означают, что значения данных не "выстраиваются" в соответствии с Пуассоном, как мы думали. Это означает, что смесь еще не однородна. Низкие р-значения не подразумевают, что теория о частицах, распределяющихся через пуассоновские процессы в смеси неверна—они в конце концов будут такими, и мы никогда в этом не сомневаемся. Мы просто проверяем, достаточно ли прошло времени.

Альтернативная форма критерия хи-квадрат, представленная в работе Айзенберга, использует следующую статистику:

$$\sum_{i=1}^N \frac{(X_i - \bar{X})^2}{\bar{X}} \quad (9)$$

что также соответствует распределению χ^2 .

Процедура тестирования та же, но есть несколько деталей, которые заслуживают упоминания. Во-первых, это интерпретация результатов. Тест, выполняемый по уравнению (8), проверяет, согласуются ли наши наблюдения с истинным средним значением λ , в то время как тест по уравнению (9) проверяет, согласуются ли наши наблюдения друг с другом. Второе различие касается степеней свободы, которые рассматриваются ниже (уравнение (9) будет иметь на одну степень свободы меньше, так как мы используем одну степень для оценки \bar{X}). Оба теста имеют свои достоинства, но необходимо быть осторожным при определении того, что использовать. Верно то, что оба будут проверять вероятность того, что набор образцов был взят из одной и той же популяции, что может означать однородность смеси (наличие разных популяций в смесителе явно означает, что он не однороден). Однако вполне возможно, что все мои образцы принадлежат к одной и той же популяции, даже если смесь не однородна. Может быть, я беру несколько образцов из "застойного" места в миксере, или, может быть, в миксере есть места, которые идентичны друг другу, но отличаются друг от друга по расположению.

Уравнение (8) будет лучше улавливать это, чем уравнение (9), потому что (8) требует не только того, чтобы образцы были взяты из идентичных популяций, но и того, чтобы каждая из этих популяций имела правильное среднее. Уравнение (9) требует только, чтобы выборки были взяты из популяций с одинаковым средним значением. Если мы уверены, что знаем истинное значение λ , мы должны использовать уравнение (8).

1.2.3 Степени свободы

Правило для степеней свободы в тесте “степени соответствия”, где используется критерий хи-квадрат, таково: $N - r - 1$, где N - количество наблюдений, а r - количество параметров, которые мы должны оценить, чтобы узнать форму теоретического распределения. Для распределения Пуассона, единственный параметр, который необходим для полного описания формы распределения это λ , таким образом степеней свободы должно быть $N - 1$ (если мы используем известную λ) или $N - 2$ (если мы используем \bar{X} вместо λ). Обратите внимание на то, что уменьшение количества образцов N ухудшает способность правильно определять моменты, когда смесь все еще неоднородна, так как N влияет на степени свободы, что, в свою очередь, влияет на критические значения хи-квадрат.

1.3 Ошибки типа I и типа II, а также определение "хорошей смеси"

В предыдущем абзаце было сделано предположение, которое должно быть сформулировано явно:

нулевая гипотеза, которую мы тестируем, заключается в том, что смесь либо однородна, либо полностью смешана. Это означает, что ошибка типа I (отклонение истинного нуля) приводит к дальнейшему смешиванию, когда мы могли бы остановиться. Ошибка типа II (принятие ложного нуля) означает, что мы делаем заключение, что смешивание полное, когда это не так. Вероятность ошибки II типа имеет обратную связь с “мощностью” статистического теста, который мы используем. Если нам нужен мощный тест (чтобы мы могли отклонить больше нулей), то мы будем делать ошибки типа II чаще. Мощность теста также связана с оптимальным размером выборки, необходимым для выполнения статистического анализа. Как уже упоминалось выше, меньшие по размеру выборки (меньшие N) означают меньшую мощность.

Это подчеркивает фундаментальный вопрос в определении состояния смеси—как мы описываем различные степени гетерогенности? Что значит сказать что-то вроде “партия смешана только на 80%”? Это то же самое, что сказать: “вероятность ошибки типа II составляет 20%”? Нет, это не так. В первом случае имеет дело с уровнем однородности внутри смеси, во втором случае имеет дело только с критическим значением.

По существу, “хорошо перемешано” и “очень, очень хорошо перемешано” будут оцениваться по-разному в первом случае, но одинаково во втором. Если вы хотите сказать: “Я правильно определяю хорошую смесь в 80% случаев”, это можно сделать, отрегулировав мощность теста, но это ничего не говорит о том, что эти 80% собой представляют (возможно, они были смешаны в два раза дольше, чем это необходимо).

Аналогично можно сказать и о выборе критического значения для отбраковки нулевой гипотезы (мы называем это “размером” теста). Для этого процесса были предложены как 5%-ные, так и 1%-ные размеры тестов, где меньший размер означает, что я буду реже отклонять нуль гипотезу. Это не то же самое, что сказать, что я буду отбраковывать только очень плохие смеси.

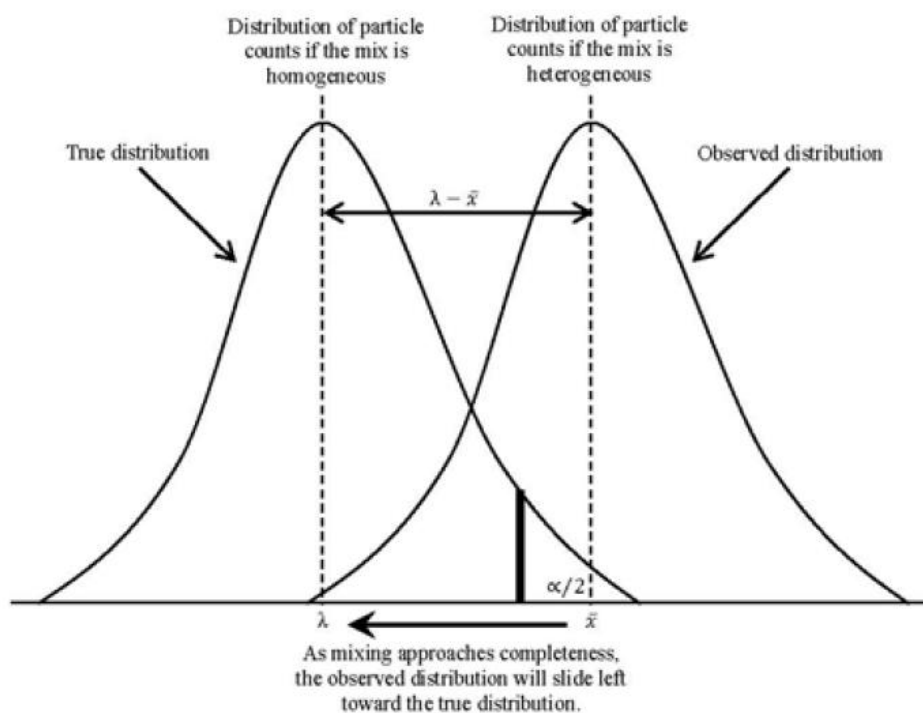
Размер теста определяет, насколько тщательно я разделяю смеси на две группы: “хорошая смесь” и “плохая смесь”. Внутри каждой группы будут различные уровни смешивания—плохие смеси и абсолютно ужасные смеси будут в одной и той же группе—и размер теста никак не влияет на возможность различить эти градации. И то, и другое будет отвергнуто одинаково.

Так в чем же здесь смысл? Определение уровня комфорта при работе с ошибками I типа и ошибками II типа является ключевым моментом. Мы должны иметь в виду, что используемая здесь статистика дает нам только ответы “да” или “нет”—они не говорят нам “насколько хорошо она смешана?”.

В результате абсолютно первостепенное значение имеет тщательное рассмотрение цены каждого типа ошибок. Можно усовершенствовать статистический тест, пока он не будет соответствовать выбранному вами критическому значению с абсолютной точностью, но этот тест будет бесполезен, если вы выберете неподходящее критическое значение.

1.4 Уточненная вероятность

Как я уже описывал ранее, при выборке из смесителя мы, по сути, строим график из полученных данных и смотрим, совпадает ли этот график с истинным распределением. Теория вероятностей и тестирование, описанные выше, могут быть уточнены с помощью простой диаграммы представленной ниже. По мере смешивания среднее число наблюдаемых частиц в образце должно приближаться к ожидаемому (истинному) среднему. По существу, наблюдаемое распределение движется к истинному распределению, как показано на рисунке ниже, где истинное среднее значение равно λ , а наблюдаемое среднее \bar{x} .



Хотя на рисунке выше подразумеваются нормально распределенные переменные, пожалуйста, имейте в виду, что истинное распределение является пуассоновским (хотя для больших λ они выглядят очень похожими). Кроме того, хотя я нарисовал наблюдаемое распределение гладким, а оно в действительности должно быть набором точек. Наконец, дисперсии каждого распределения кажутся идентичными—деталь, которая может не соответствовать действительности (наблюдаемое распределение может быть шире истинного распределения, например). Однако эти предположения делаются только для того, чтобы уменьшить беспорядок на рисунке.

Ослабление этих предположений не подрывает результатов или правильность последующих рассуждений.

Исходя из ситуации, изображенной выше, мы бы отвергли нулевую гипотезу о том, что смесь является полной (или однородной), поскольку наблюдаемое среднее значение \bar{x} лежит справа от критического значения (темная линия, отделяющая λ от области, помеченной меткой $\alpha/2$). Если \bar{x} находится дальше от λ , чем критическое значение, то у нас есть уверенность, что \bar{x} будет настолько редким случаем, что при истинной нуль гипотезе мы отбрасываем эту нуль гипотезу.

Как ошибки типа I и типа II влияют на наши рассуждения? Рассмотрим следующие четыре возможных результата статистического теста, основанного на приведенном выше рисунке:

1. Распределения различны, и я получаю значение для \bar{x} , которое находится справа от критического значения, и в этом случае я отклоняю нуль гипотезу(правильно).

2. Распределения различны, и я получаю значение для \bar{x} , которое находится слева от критического значения, и в этом случае я принимаю нуль гипотезу (неправильно). Ошибка ВТОРОГО типа.
3. Распределения одинаковы, и я получаю значение для \bar{x} , которое находится справа от критического значения, и в этом случае я отклоняю нуль гипотезу (неправильно). Ошибка I типа.
4. Распределения одинаковы, и я получаю значение для \bar{x} , которое находится слева критического значения, и в этом случае я принимаю нуль гипотезу (правильно).

Очевидно, что корректировка критического значения повлияет на вероятность ошибок, указанных в пунктах #2 и #3, а также на фактическое положение каждого распределения. Если бы я выбрал очень маленькое значение для α такое, чтобы уверенно отвергнуть нуль гипотезу, тогда критическое значение (темная линия) должна переместиться вправо—это будет означать, что группа, которую я считаю хорошими смесями, начнет включать все больше и больше плохих смесей. Если переместить темную линию, очерчивающую критическое значение на этой диаграмме, вправо, до положения, пока она не окажется справа от \bar{x} , то в этом случае я бы принял вышеприведенное наблюдаемое распределение как статистически равное истинному распределению, и, таким образом, пришел бы к выводу, что смесь хороша, хотя это и не так. Эта цифра и понимание взаимосвязи между формой и положением распределений относительно друг друга помогут нам ответить на вопросы, касающиеся таких вещей, как оптимальный размер выборки и количество частиц. Мораль всего, что следует далее, заключается в том, насколько важно иметь глубокое понимание того, что на самом деле означают тип I и тип II ошибок с точки зрения процесса смешивания кормов.

2. Сколько образцов должно быть взято из смеси?

2.1 Важность размера выборки N

Этот вопрос касается "размера выборки" в статистическом смысле - речь идет о количестве наблюдений, которое я называю N. Размер выборки N не совпадает с "размером каждого образца", который соответствует количеству корма (в граммах), взятого из смесителя для анализа. Как правило, благодаря таким вещам, как Закон больших чисел (среднее значение выборки приближается к истинному среднему значению популяции по мере роста числа наблюдений) и Теореме о Центральном пределе (средние значения выборки будут нормально распределены для больших размеров выборки), большие размеры выборки всегда желательны. К сожалению, не существует ни одного правила или общепринятого определения понятия "достаточно большой". Этот вопрос усложняется еще и тем, что распределение Пуассона измеряет как успехи, так и неудачи—так что размер образцов и количество необходимых наблюдений трудно определить.

Например, если я бросаю монету 10 раз и выпадают 3 решки, каков размер выборки?

Это 10, а не 3. В этом смысле "размер выборки" следует интерпретировать как "число испытаний", а не "количество успехов". В контексте тестирования корма размер выборки определяется количеством образцов, взятых из партии для тестирования. Количество подсчитанных частиц не является размером выборки.

Рассмотрим другой пример, когда я измеряю рост людей в популяции и хочу знать, нормально ли распределены значения роста. Я хватаю человека, беру линейку и записываю его рост. Теперь у меня есть одно наблюдение, так что $N = 1$, используя это измерение высоты, я ставлю точку на своем графике нормального распределения и я понимаю, что не в состоянии сказать что-нибудь о форме наблюдаемого распределения - у меня есть только одна точка, а мне нужно знать, выглядит ли распределение как кривая в форме колокола. Измерение роста второго человека поможет—использование более точной линейки не поможет. Увеличение количества образцов - это первое, увеличение количества трейсеров в партии - второе.

Статистические тесты, которые мы используем, аналогичны сравнению графика данных с графиком правильного распределения Пуассона и сравнению их форм.

В результате, если у меня есть только несколько точек, я не смогу однозначно сделать вывод о форме наблюдаемого распределения, и это снизит мою уверенность в результатах теста. Однако,

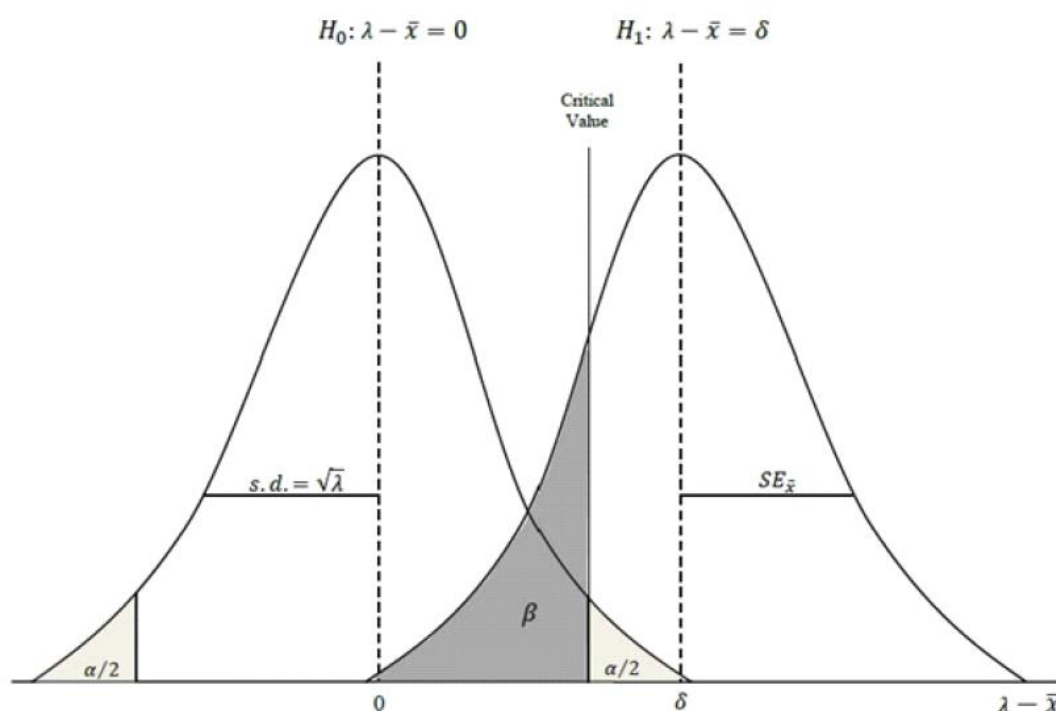
если у меня есть 50 наблюдений, то я в состоянии сказать, как выглядит их график распределения. Сколько их достаточно?

Опять же, нет ни одного правильного способа ответить на этот вопрос, но мне было бы очень неудобно иметь менее 10 наблюдений, 20 было бы хорошо, 30 было бы роскошно—больше этого не нужно.

3 Уравнение Лера и размер выборки N

Формализованный метод определения подходящего размера выборки для статистического тестирования был предложен Лером (Lehr, 1992). Требуется некоторая терминология и обозначения, которые я буду приводить здесь. Следующий рисунок предназначен для того, чтобы помочь в понимании этого раздела.

Обратите внимание: хотя рисунок ниже очень похож на рисунок в разделе 1.4, они очень сильно отличаются. На предыдущем рисунке показано среднее значение по оси x. На рисунке ниже показана разница средних значений по оси x.



Как и в случае с первым рисунком в этой статье, они кажутся нормальными распределениями, но они будут пуассоновскими. Распределение справа—это наблюдаемые данные, а распределение слева - истинное распределение. Обозначения, используемые здесь являются стандартом. SE_x стандартная ошибка для наблюдаемого распределения, а поскольку это пуассоновское распределение, $\sqrt{\lambda}$ - стандартное отклонение для истинного распределения (на рисунке они кажутся равными, но это не обязательно), \bar{x} - среднее значение выборки, λ - истинное среднее. Индекс 0 указывает на нуль гипотезу, 1 - на альтернативу, а H_0 - нулевая гипотеза (однородная смесь), которая требует, чтобы выборочное среднее было таким же, как истинное среднее, $\lambda - \bar{x} = 0$. (Я нарисовал это так, что наблюдаемое распределение находится справа от нуля ($\bar{x} > \lambda$), но вместо этого может быть верно обратное.) Вероятность ошибки типа I равна α (размер теста), а вероятность ошибки типа II равна β (темная область). Мощность теста, таким образом, равна $1 - \beta$. Обратите внимание, что здесь я предполагаю 2-сторонний тест, поэтому каждый хвост помечен $\alpha/2$. Здесь мы будем проверять нулевую гипотезу о том, что разница в средних значениях равна нулю. Это согласуется с тестом, который мы используем при смешивании корма—мы хотим знать, совпадает ли фактическое среднее число частиц с предсказанным теоретическим значением. Обратите внимание, что нулевое распределение определяет размер теста α , а альтернативное

определяет мощность теста $1 - \beta$. Если бы я выбрал α такой, чтобы уверенно отвергнуть нуль гипотезу (α было бы —1% или 0,1%, например), тогда критическое значение (обозначенное на рисунке) сдвинулось бы вправо, а β выросло — что значило бы, что моя способность правильно идентифицировать неполную смесь падает. Я буду одобрять более неполные смеси по мере приближения α к нулю.

Таким образом, более низкая ошибка I типа (малая α) подразумевает более высокую ошибку II типа (большая β), при прочих равных условиях.

Формально расстояние между критическим значением и нулем H_0 равно $0 - Z_{1-\alpha/2}\sqrt{\lambda}$, а расстояние между альтернативой H_1 и критическим значением равно $\delta - Z_{1-\beta}SE_{\bar{x}}$, где значение Z взято из стандартной таблицы нормального распределения. Здесь используется буква Z потому, что мы имеем дело с нормальными распределениями. Я буду исходить из предположения, что $\sqrt{\lambda} = SE_{\bar{x}}$ только для упрощения. Это предположение может быть смягчено без ущерба для результатов.

Используя приведенную выше информацию, Лер вывел следующую формулу для размера выборки:

$$N = \frac{2(Z_{1-\alpha/2} + Z_{1-\beta})^2}{\left(\frac{\mu_0 - \mu_1}{\sigma}\right)^2} \quad (10)$$

Числитель для $\alpha = 5\%$ и $\beta = 20\%$ равен 15,68, который я в дальнейшем округляю до 16. Это число вычисляется с использованием табличных значений нормального распределения для выбранных значений α и β .

Теперь мы можем перейти к распределению Пуассона. Известно, что если случайная величина X имеет распределение Пуассона со средним значением λ , то \sqrt{X} распределена нормально со средним значением $\sqrt{\lambda}$ и дисперсией $\sigma^2 = \frac{1}{4}$. Это результат так называемого "дисперсионного стабилизирующего преобразования". Этот факт и уравнение (10) подразумевают, что правило определения размера выборки для распределения Пуассона с использованием $\alpha = 5\%$ и $\beta = 20\%$ таково:

$$N = \frac{4}{(\sqrt{\lambda_0} - \sqrt{\lambda_1})^2} \quad (11)$$

Если отбросить всю математику, как мы можем использовать эту формулу? Первоначально у нас возникло желание утверждать, что λ_0 и λ_1 равны, поэтому знаменатель этой дроби равен нулю. Однако λ_0 и λ_1 равны только тогда, когда смесь совершенно однородна. Если я добавил частицы микротрейсера всего несколько минут назад, у меня нет оснований ожидать, что среднее значение выборки будет равно истинному среднему. Таким образом, эта формула в сочетании с моим пониманием того, как происходит смешивание, может быть использована для определения того, сколько образцов должно быть взято в зависимости от времени. По мере того, как значения λ_0 и λ_1 сближаются (то есть по мере того, как смешивание приближается к завершению), знаменатель становится меньше, и поэтому мой оптимальный N должен стать больше. Таким образом, мне нужно брать больше образцов по мере смешивания. Сколько их будет, зависит от моего ожидания того, как наблюдаемое среднее будет изменяться с течением времени.

Например, рассмотрим случай, когда $\lambda_0 = 50$ и я смешиваю в течение определенного периода и считаю, что смесь готова на 80%. Если это было бы так, то $\lambda_1 = 40$ и эту гипотезу стоит проверить, чтобы подтвердить свои убеждения о статусе смеси². Учитывая приведенную выше формулу, я бы проверил по крайней мере $N = 7$ образцов смеси.

Если бы вместо этого я считал, что смесь готова на 90%, то при $\lambda_1 = 45$ потребовалось бы проверить $N = 30$ образцов смеси.

² Обратите внимание, что это не обязательно единственная интерпретация утверждения "смесь на 80% завершена".

4. Сколько частиц следует добавить в смесь?

Из-за таких вещей, как погрешность измерения, точное количество частиц Микротрейсера не известно с идеальной точностью. Если мы говорим 25 000 частиц на грамм, мы на самом деле имеем в виду что-то близкое к 25 000, но, как и любая другая оценка, лучшее, что мы можем сделать, - это доверительный интервал. Эта неопределенность означает, что мы на самом деле не знаем точно, каково истинное значение λ , но, как мы видели, у нас есть методы построения доверительных интервалов вокруг λ на основе выборочных данных.

Неопределенность в отношении количества частиц осложняет дальнейшую дискуссию, особенно в контексте доклада ТНО. Например, рассмотрим ситуацию, когда мы определили, что увеличение до $\lambda = 400$ улучшит наши результаты на 10% (мысль, к которой мы вернемся в ближайшее время). Для того, чтобы сделать это утверждение (или для того, чтобы выполнить эту политику), нам потребовалась бы значительная точность относительно нашего начального количества частиц микротрейсера, которые будут добавлены в смеситель.

Знание того, что $\lambda = 400$ на 10% лучше, чем $\lambda = 300$, ценно для меня только в том случае, если я действительно могу отличить $\lambda = 400$ от $\lambda = 300$. Если ошибка измерения превышает этот уровень точности в моем измерении начального количества частиц, то статистическое превосходство $\lambda = 400$ не имеет значения.

Суть в следующем: незнание в совершенстве количества частиц на грамм смеси означает, что предельные корректировки λ будут проблематичными.

4.1 (Не)важность ожидаемого количества частиц

Предположим, мы смешиваем 3000 кг корма, к которому добавляем частицы микротрейсера (25 000/грамм) из расчета 50 г на 1000 кг корма. Количество частиц, которое мы ожидаем найти в образце 80г. смеси, известно в момент добавления микротрейсеров. При соотношении, используемом здесь, мы ожидали бы увидеть 1,25 частиц на грамм корма или 100 частиц в нашем образце весом 80 г, если бы смесь была совершенно однородной (то есть идеально Пуассоновской)³. Это λ , истинное среднее значение популяции — которое находится там, где располагается центр фактического распределения Пуассона для этой партии. Если я хочу λ большего размера, есть два способа, которыми я могу действовать: я могу либо взять большие образцы из смесителя (160 г вместо 80 г), либо я могу добавить больше частиц в смесь. Последний вариант (добавление большего количества частиц) повлияет на то, как я интерпретирую результаты статистического тестирования. Причина заключается в том, что каждый аспект формы распределения Пуассона зависит от значения λ . Если я добавляю больше частиц, я увеличиваю λ , что, в свою очередь, увеличивает дисперсию, которая делает распределение жирнее. Большая дисперсия увеличит размер хвостов моего распределения, что увеличит вероятность ошибок типа I, если я не скорректирую свое критическое значение (обычно обозначается как α) для компенсации.

4.1.1 Быть введенным в заблуждение CV

Как показывает формула (2), коэффициент вариации является функцией λ . А именно, CV имеет обратную зависимость от λ . Добавление большего количества частиц в смесь уменьшит значение CV. Как и в нашем предыдущем обсуждении, интерпретация того, что это означает, является ключевой. CV говорит мне о том, как мои измерения варьируются по отношению к среднему значению (думайте об этом как об отклонении в процентах от среднего). Если повторно использовать пример с измерением роста, приведенный выше, то CV сообщило бы мне, как мои измерения роста варьируются для низкорослых людей по сравнению с вариациями измерений для высоких людей.

Допустим, что я ошибаюсь при измерении в среднем на 2 дюйма для человека ростом 6 футов, то я

³ Мне также нужно предположить, что ошибка подсчета частиц равна нулю и никакие частицы не теряются— предположения, которые являются смелыми в некоторых ситуациях.

буду ошибаться не более чем на 2 дюйма, когда я измеряю человека ростом 5 футов. Если бы было наоборот, я бы хуже, в процентах, измерял низкорослых людей, чем высоких, что потребовало бы некоторого теоретического обоснования. Мы можем интерпретировать меньшее CV в этом примере по существу, как результат использования более точной линейки. Когда дело доходит до микротрейсеров в корме, ход мыслей аналогичен. Если я ожидаю только 4 частицы в образце, и я упускаю одну частицу из-за ошибки измерения, значит, я упустил 25% частиц. Увеличение размера образца позволит мне упустить такое же количество частиц при уменьшении процентного отклонения.

Однако в контексте тестирования смеси кормов мы не сильно озабочены улучшением CV. Правда, более низкий CV подразумевает меньший процент вариации в моем подсчете микротрейсерных частиц для каждого образца, что согласуется с аналогией "лучшей линейки". Кроме того, использование лучшей линейки позволит мне более точно разместить результаты измерений, которые я только что сделал, на моем графике распределения Пуассона. Однако, как мы уже видели, точное расположение каждой точки не так важно, как то, насколько точно расположение точек по форме совпадает с теоретическим распределением Пуассона. Улучшение CV нам в этом не поможет. Хуже того, влюбленность в низкое CV может дать нам ложное чувство уверенности в наших результатах.

4.2 Размер выборки и критерий Пуассона

Как мы знаем, и как обсуждалось выше, я могу использовать выборочные наблюдения, чтобы оценить, согласуется ли форма распределения с истинным распределением Пуассона. Чтобы это сделать, используем формулу (1) и статистические таблицы, по которым находим критические значения случайных величин распределения Пуассона. Придерживаясь идеи о том, что увеличение количества частиц микротрейсера в партии помогает более точно позиционировать данные на моем графике, следовательно, позволяет улучшать и результаты теста Пуассона. Это технически верно, но, как и прежде, мы должны быть осторожны, чтобы тщательно интерпретировать улучшение.

Из того, что мы знаем о распределении Пуассона, можно сказать, что значения λ , которые больше 10, достаточны для статистической проверки. Проблема малого λ в отношении критерия Пуассона для формы распределения заключается в следующем: если λ очень близко к нулю, то я могу обнаружить, что левый хвост моего распределения Пуассона обрублен, что может затруднить мне правильное построение формы распределения.

Эта логика мотивирует наш выбор "Больших" значений. Большие значения λ не вредят, но экономическая выгода асимптотически приближаться к нулю. Это означает, что если выгода превышает стоимость теста, то стоимость дополнительного количества частиц равна нулю. Поскольку опять же нет единого правила, которому можно следовать, и основываясь на том, что мы знаем об асимптотике и распределении Пуассона, я предлагаю $20 \leq \lambda \leq 100$.

4.3 Размер выборки и тест χ^2

Критерий хи-квадрат, описанный выше, по существу проверяет местоположение среднего значения моего наблюдаемого распределения и сравнивает его с местоположением среднего значения для истинного распределения. Если я добавлю в корм такое количество микротрейсеров, при котором ожидаю получить 100 частиц в каждом образце весом 80 г, то расположение среднего истинного распределения будет равно 100. Если я найду только 62 частицы в образце, то тест хи-квадрат покажет мне вероятность увидеть только 62 частицы трейсера, учитывая, что я ожидал увидеть 100. Если я добавлю дополнительные частицы, чтобы увеличить λ , то я смещаю оба распределения— истинное и наблюдаемое—вправо (см. рисунок выше). Расстояние между этими двумя средними не изменится. Однако, поскольку дисперсия распределения Пуассона увеличивается с ростом λ , я ожидаю увидеть больше наблюдений вдали от среднего—это может заставить меня сделать заключение, что мое наблюдаемое распределение находится не в том месте,—поэтому я буду продолжать смешивать без необходимости. Таким образом, если я увеличиваю количество трейсеров, я должен также ужесточить стандарты, по которым я оцениваю свои результаты—

увеличение числа трейсеров должно сопровождаться ужесточением критических значений. Это можно рассматривать как улучшение в нашем тестировании — и это технически так, поскольку мы будем более точны в наших отказах, но это улучшение в статистическом тестировании, вероятно, не приведет к улучшению качества корма. Увеличение количества частиц в корме поможет нам определить, насколько точно среднее значение истинного распределения отличается от среднего значения наблюдаемого распределения. Однако это преимущество применимо только в тех случаях, когда мы очень близки к критическому значению.

Увеличение λ ничего не делает в контексте помощи в определении соответствующего критического значения, что, как мы видели, является фундаментальным вопросом. Знание того, что я нахожусь в пределах ангстрема критического значения, важно только в том случае, если критическое значение имеет важное значение. Связь между размером выборки и критическими значениями будет рассмотрена в следующем разделе.

5 Насколько большой должна быть выборка?

Чтобы избежать путаницы между "размером образца" и размером выборки N . Я буду говорить "масса образца". Я предполагаю, что образцы измеряются в граммах. Это предположение, хотя и кажется безобидным, на самом деле может создать проблемы, если масса некоторых ингредиентов в корме сильно отличается от массы других в пересчете на единицу объема. Альтернативным методом было бы определение размера образца в единицах объема, но это может так же создать проблемы по другим причинам (данный вариант, вероятно, лучше всего подходит для жидкостей). Заметим, что масса взятого образца тесно связана с количеством трейсеров, которые мы ожидаем найти — большой по размеру совок должен содержать больше трейсеров. Зная это, мы должны установить массу образца такой, чтобы исключить риск попасть в затруднительную ситуацию, заигрывая с маленькими значениями для λ . Рекомендации, приведенные в предыдущем разделе, помогут нам в этом.

Еще следует учитывать, что измерение массы образца само по себе не является точной наукой — мы будем иметь среднее значение и дисперсию для каждой массы, хотя распределение будет приблизительно нормальным (не пуассоновским), пока мы берем разумное количество образцов. То, что мы знаем о таких вещах, как CV, это то, что более массивные выборки будут менее изменчивы в процентном выражении. Представьте шкалу имеющую точность до 0,1 грамма. В средней школе мы учили, что если это так, то 10,1 г и 10,2 г должны считаться одной и той же массой исходя из разряда значимых цифр. Отсюда следует, что если масса образцов будет достаточно большой, то это приведет к более точному результату. Однако чрезмерно массивные образцы могут снизить точность. Интуитивно мы можем представить себе предельный случай, когда размер выборки настолько велик, что равен всему объему смесителя. В этом случае мы можем точно подсчитать количество микротрейсеров, но не будем иметь информацию о том, как микротрейсеры распределены по объему партии корма. Конечно, предельный случай в противоположном направлении не лучше — взятие таких крошечных образцов, в которых нет частиц, явно не даст надежных результатов. Тем не менее, когда дело доходит до определения массы образца, который нужно взять, наша единственная забота должна заключаться в том, что бы λ являлась приемлемо большой. Кроме того, до тех пор, пока масса образца намного меньше массы всей смешиваемой партии, любой удобный размер должен быть достаточным.

6. Откуда следует брать образцы?

Этот вопрос был затронут в оригинальной статье Айзенберга, хотя ответ не был формализован. Проблема заключается в том, что микротрейсеры изначально добавляются в одно место (или, возможно, в несколько мест) в смесителе. В результате образцы, взятые из областей, расположенных вблизи этой точки, будут иметь иное распределение микротрейсеров, чем образцы, взятые из более отдаленных областей в смесителе, по крайней мере, если время смешивания было коротким и полного смешивания не было достигнуто. Это может привести меня

к выводу, что смесь лучше, чем она есть на самом деле, если я случайно возьму образцы из того места в смесителе, где частицы совпадают с ожидаемыми мной значениями.

Существует несколько решений этой проблемы. Первое заключается в том, чтобы взять образцы одновременно из нескольких мест в смеси. В качестве альтернативы я мог бы взять несколько образцов из одного и того же места, но в разные временные интервалы процесса смешивания. На самом деле эти два метода могут привести к разным результатам. Например, если тяжелые частицы тонут, а более легкие поднимаются в смеси, то отбор проб из одного места в смесителе многократно с течением времени может дать результат что смесь однородна, когда на самом деле имеет значительную неоднородность.

Вполне возможно, так же, что определенная область в смесителе перемешивается не так хорошо, как другие, и если мне не повезет и я возьму образец из этого места, мои результаты не будут точно отражать степень однородности. Из этого следует, что сочетание того и другого—пробы из нескольких мест, взятые в разное время — дадут наилучшие результаты.

Рассмотрим пример, где микротрейсеры добавляются в точку А, которая равно удалена от какой-то другой точки В. Я буду брать образцы весом 100 г в разные временные интервалы, из места точно посередине между точками А и В. В промежутках между заборами образцов частицы трейсера слегка мигрируют в кормовой смеси. В результате число частиц, наблюдаемых в выбранном мной месте, должно со временем изменяться. Первоначально я должен увидеть меньше частиц, чем ожидалось (так как я далек от точки А, где были добавлены частицы), и это число должно приближаться к ожидаемому мной значению по мере продолжения смешивания. Если мои наблюдения наконец совпадут с моими ожиданиями, идеальная ли это будет смесь? Поскольку я никогда не брал пробы из точки В, я технически не знаю, как выглядит смесь в этом месте. Увеличение числа мест выборки решает эту проблему.

Следует также отметить, что физический мир иногда не согласуется с миром, который мы описываем на бумаге. Например, микротрейсеры могут покрываться слизью по мере увеличения времени смешивания, что затрудняет их обнаружение. Это означает, что процесс приближения результатов к ожидаемым значениям не идеален (он может даже не быть монотонным), и выборка из одного места с течением времени может быть не идеальной.

Так как же определить место отбора проб? Было бы трудно оправдать отсутствие более чем одного места и более чем одного времени отбора проб.

Имейте в виду, что реальная проблема здесь заключается в размере выборки N . Если мы желаем, чтобы $N = 10$, то, возможно, 2 места в 5-ти разных временных интервалах вполне могут подойти. Как эти результаты будут отличаться от взятия из 5 мест в 2-ух разных временных интервалах? Как я уже говорил выше, первый вариант может быть более уязвим к таким вещам, как застойные места в смесителе, в то время как другой может быть более уязвим к таким вещам, как плавающие и тонущие частицы.

В соответствии с общей темой обсуждения решение этого вопроса должно основываться на учете затрат и выгод.

Я не могу не подчеркнуть, что время выборки должно быть выбрано тщательно, чтобы избежать ошибок в статистических тестах. Если я намерен использовать образцы выборки из разных временных интервалов, то я должен быть уверен, что все образцы, взятые для теста, согласуются с альтернативной гипотезой. Например, я не могу взять образец через 10 секунд после начала смешивания, а затем еще один образец через 20 минут смешивания, если я намерен использовать эти образцы в одном и том же тесте гипотезы. Нуль гипотеза в обоих временных периодах будет совпадать (полное смешивание), но альтернативная гипотеза (λ_1) будет сильно отличаться. (См. обсуждение уравнения (11) выше.) Я могу использовать эти наблюдения для оценки сходимости среднего, но я не могу использовать их для проверки положения среднего в том же тесте. Это согласуется с идеей о том, что должно быть принято решение о том, что является желательным - положение среднего или сходимость. Имеет ли это смысл, если возьму пробы в двух разных временных интервалах—если в обеих временных интервалах у меня было одинаковое ожидание относительно значения λ_1 ? Это является основой для моей рекомендации о многократном отборе проб.

7. Сколько суб-образцов должно быть взято из каждого образца?

Использование нескольких подвыборок из одного образца может быть проблематичным в том смысле, что наблюдаемое количество частиц в подвыборке А не будет независимым от количества частиц в подвыборке В. Например, если исходный образец была взят из части смеси, в которой наблюдается высокая концентрация частиц микротрейсера, то можно было бы ожидать, что обе подвыборки будут иметь так же большое количество частиц. Если мы нарушаем допущение независимости, мы подрываем достоверность наших статистических данных—например, тест хи-квадрат может больше не быть жизнеспособным. Тем не менее, разделение выборки на подвыборки все еще может иметь место, но значения подвыборок должны быть каким-то образом объединены (например, усреднены) вместо того, чтобы использоваться отдельно как два разных наблюдения. Смысл, лежащий в основе этого риска, связана с обсуждавшейся в предыдущем разделе проблеме, связанной со временем и местом отбора проб. Разделение пробы на две подпробы эквивалентно взятию двух проб из одного и того же места в одно и то же время. Так как микротрейсеры не успевают перемещаться между интервалами отбора проб и расстояние между точками взятия проб равно нулю, т.е. две подпробы дают не больше информации об однородности смеси, чем одна проба, взятая в одно и то же время и в одном и том же месте.

8. Отчет TNO

Отчет TNO начинается с нескольких примеров, которые вызывают несогласие с принятым компромиссом между ошибками типа I и типа II. В частности, использование 1% - ного доверительного интервала, проверяющего нулевую гипотезу о том, что смесь хорошая. Результаты в этих моделях были приемлемы для ошибок типа I, но неприемлемы для ошибок типа II—они ошибочно принимали слишком много плохих смесей за хорошие.

Стандартное решение состоит в том, чтобы переместить критические значения ближе к среднему значению (больше α), что приведет к тому, что в хвостах будет зафиксировано больше наблюдений. Так как больше наблюдений теперь находятся дальше от среднего, чем критическое значение, соответственно будет больше отказов от нулевой гипотезы (что смесь хороша) Объявление смеси хорошей реже означает меньше ложных "хороших"—но это также может означать меньше истинных "хороших". Это именно то, что они обнаружили—увеличение α до 5% уменьшило ошибки типа II до приемлемого уровня, но слишком многим было пожертвовано с точки зрения ошибок типа I—они ошибочно приняли слишком много хороших партий за плохие.

Проблема, с которой сталкиваются авторы, заключается в следующем: поскольку этот компромисс между ошибками типа I и ошибками типа II влияет на каждую из их имитационных моделей по-разному, исходя из параметров каждой (количество частиц и т. д.), становится трудно сравнивать два разных метода оценки качества смеси. Одно количество частиц или размер выборки может работать лучше для ошибок типа I, а другое лучше для ошибок типа II. Как нам выбрать, что важнее?⁴ Чтобы ответить на этот вопрос, отчет TNO использует данные компьютерного моделирование и сравнения сотен моделей с различными параметрами⁵.

Они также варьировали такие вещи, как погрешность измерения и уровень неоднородности смеси. На основе результатов моделирования было определено наилучшее количество частиц, размер выборки и использование суб-проб. Прежде чем я продолжу, я хочу сделать предупреждение о численных методах: моделирование может быть очень полезным, но я думаю, что оно имеет тенденцию скрывать смысл результатов под массой очень привлекательных данных. По существу, они смещают фокус с понимания на интерпретацию данных, что может быть рискованно. Тем не менее, как мы увидим далее, эти результаты очень полезны для понимания рассматриваемой проблемы.

Сравнение различных моделей в отчете TNO осуществляется следующим образом. "Хорошей" смесью признается смесь, в которой распределение частиц является достаточно «худым» (низкая

⁴ В этот момент я бы глубоко задумался о значении и экономическом воздействии каждого типа ошибок. Только глубокое понимание производимого продукта, затрат и выгод от каждого типа ошибки и общего контекста проблемы может дать ответ на этот вопрос

⁵ Параметрами были: λ (от 75 до 750), N (10 или 20) и количество суб-образцов (1 или 2).

дисперсия), так что 68,2% наблюдений попадают в предел отклонения от среднего равный $\pm 4\%$ ⁶. Авторы установили стандарт принятия хороших смесей в каждой тестируемой модели равной 80%, регулируя размер α до тех пор, пока 80% смесей, которые действительно хорошие, не будут помечены как хорошие. Если модель с определенными значениями параметров правильно принимает только 70% хороших смесей, то авторы немного расширяют доверительный интервал (меньше α), чтобы уменьшить количество отказов и увеличить коэффициент принятия до 80%. После такой нормализации каждая модель будет пропускать одинаковое количество хороших смесей. Поскольку нормализованные модели работают одинаково для хороших смесей, каждая модель оценивается на основе того, сколько плохих смесей проходит—очевидно, чем меньше, тем лучше⁷. "Плохой" смесью признается смесь, к которой распределение частиц является достаточно «жирным» (высокая дисперсия), так что 68,2% наблюдений попадают в предел отклонения от среднего равный $\pm 10\%$. Поскольку по определению "хорошие" смеси имеют отклонение $SE \leq 4\%$ от среднего значения, то по существу мы проверяем, попадает ли стандартное отклонение выборки, из которой наши данные получены (с известным λ) в интервал меньше или равно $0,04\lambda$. Если это так, то нам нужно определить эти 80% результатов в соответствии с нашими правилами нормализации. Таким образом, этот тест касается доверительного интервала вокруг стандартной ошибки для распределения, из которого мы извлекли наши данные. Формула для доверительного интервала вокруг расчетной стандартной ошибки в этой ситуации однозначна:

$$SE \sqrt{\frac{N-1}{\chi^2(\alpha/2; N-1)}} \leq \sigma \leq SE \sqrt{\frac{N-1}{\chi^2(1-\alpha/2; N-1)}} \quad (12)$$

где $\chi^2(p; n)$ значение хи-квадрат для уровня значимости p и степеней свободы n .

Напомним, что для распределения Пуассона $\sigma = \sqrt{\lambda}$. Также напомним, что нашим стандартом для хорошей смеси является условие, что $SE = 0,04\lambda$. Делая соответствующие замены в приведенном выше уравнении, мы получаем следующее:

$$0,04\lambda \sqrt{\frac{N-1}{\chi^2(\alpha/2; N-1)}} \leq \sqrt{\lambda} \leq 0,04\lambda \sqrt{\frac{N-1}{\chi^2(1-\alpha/2; N-1)}} \quad (13)$$

Используя это уравнение, мы можем вычислить значение λ , как для нижней, так и для верхней границы при заданных N и α . Например, чтобы найти нижнюю границу:

$$0,04\lambda \sqrt{\frac{N-1}{\chi^2(\alpha/2; N-1)}} = \sqrt{\lambda} \quad (14)$$

⁶ Вот откуда берутся 68,2%: они определяют разброс неоднородности как $f_{nu} = \frac{SE_{nu}}{\lambda}$ и утверждают, что произвольное значение $f_{nu} = 4\%$ - это определение "хорошей" смеси. Из этого следует, что для хорошего смешивания, $SE_{nu} = f_{nu} \cdot \lambda = 0,04 \cdot \lambda$. Для нормального распределения (которое они предполагают имеет место быть) 68,2% наблюдений попадают в интервал $[\mu - SE, \mu + SE]$. Это означает, что 68,2% наблюдений в нашем случае попадают в интервал $[\lambda - .04\lambda, \lambda + .04\lambda] = [.96\lambda, 1.04\lambda]$. То есть 68,2 % попадают в диапазон $\pm 4\%$ от среднего.

⁷ Выражаясь техническим языком: "Хорошая" смесь определяется как смесь, которая имеет нормальное распределение со стандартной ошибкой, равной $0,04 \cdot \mu$. Зафиксируйте вероятность ошибки типа I на уровне 20%, а затем минимизируйте вероятность ошибки типа II.

что для $N = 20$ и $\alpha = 20\%$ дает нам следующее:

$$\underline{\lambda}(N; \alpha) = \underline{\lambda}(20; 0.2) \approx 895$$

Аналогично для верхней границы, используя те же значения для N и α , получаем:

$$\overline{\lambda}(20; 0.2) \approx 383$$

Кажется логичным, что большее значение λ определило нижнюю границу стандартного отклонения, поскольку большее количество частиц не может повредить моей оценке. Среднее из них равно 639, что является значением λ , которое дает наибольшую вероятность того, что стандартное отклонение находится в центре доверительного интервала. Таким образом, $\lambda = 639$ дает наибольшую вероятность правильной идентификации смеси с $SE=0.04\lambda$, по крайней мере в 80% случаев. Эти цифры для λ также согласуются с теми, которые были предложены в отчете TNO.

А как насчет отказа от плохих смесей? Согласно документу, TNO, плохая смесь означает, что $SE \geq 0.10\lambda$ для базового распределения, из которой мы взяли наши образцы. Используя значение λ , вычисленное выше, и $N = 20$, мы могли бы использовать аналогичный метод для определения шансов правильного определения плохой смеси. Имейте в виду, что это не совсем то, как была структурирована статья TNO, но процесс, проиллюстрированный здесь, будет определять последующие аргументы.

Есть ли какие-то проблемы с этой методологией? Вроде бы. Одна из основных проблем заключается в том, что оценка λ критически зависит от определения "хорошей" смеси. Например, если я изменю стандарт для хорошей смеси с $\pm 4\%$ на $\pm 3\%$, тогда количество частиц, необходимых для одного и того же уровня достоверности, удвоится.

Это важно, потому что стандарт 4% является довольно произвольным. В результате вычисленное значение λ также должно считаться произвольным. Это означает, что наилучшее значение λ не должно быть определено путем моделирования, а только исходя из контекста, в котором будет использоваться смесь.

Какой уровень неоднородности допустим в смеси? Используя те же расчеты и предположения, что и выше, если нас устраивает 10% неоднородность (как определено в статье TNO), мы получаем $\lambda = 100$. Это именно то значение, которое было предложено в оригинальной статье Айзенберга 30 лет назад. (Обратите внимание, что масса образцов явно не определена в отчете TNO, таким образом, ожидаемое количество частиц на образец требует здесь некоторой интерпретации.)

Есть и другие примеры подобных моментов в статье, такие как предположение об ошибке измерения и произвольный стандарт 80% для принятия хороших смесей. Ослабление (или усиление) любого из этих предположений может значительно изменить наш результат, что подразумевает для меня, то что мы не должны сильно верить в наши численные результаты, если у нас нет огромной веры в наши предположения. Ни одно из упомянутых здесь предположений не кажется достойным этой веры.

Следует также отметить, что отчет TNO не включает в свои модели многие детали, упомянутые в предыдущих разделах настоящего документа. Например, неопределенность относительно того, сколько именно частиц мы добавляем, коррелированная со временем проблема покрытия слизью микротрейсеров и разнообразие в размере и типе частиц корма, которые могут привести к всплыванию или погружению материалов. Я не утверждаю, что они заслуживают большего внимания, чем те факторы, которые рассматриваются в докладе TNO, я просто пытаюсь подчеркнуть деликатность численных методов в свете столь многих неизмеримых моментов процесса смешивания кормов.

Можно ли доверять результатам TNO? Да, можно. Являются ли результаты TNO абсолютным определением оптимальной процедуры? Нет, это не так—то, что оптимально, зависит от ситуации.

8.1 Пример

Это простой пример, призванный проиллюстрировать важность понимания того, как ошибки I типа и II типа влияют на процесс тестирования. Предположим, что у нас есть стандарт для ошибки I типа равный 5% и стандарт для ошибки II типа равный 20% (в соответствии с вышеизложенным). Учитывая эти стандарты, я хочу оценить две разные процедуры:

1. Случай, когда $N = 20$ и λ изменяется для достижения этой цели;
2. Случай, когда $\lambda = 100$ и N изменяется для достижения этой цели.

По существу, я спрашиваю, насколько далеко от истинного среднего может быть наблюдаемое среднее, если я должен обнаружить эту разницу на заявленных уровнях доверия. Я могу использовать уравнение (11), чтобы ответить на этот вопрос. Подставим в формулу значения $N = 20$, $\lambda = 100$ и мои выбранные критические значения 5% и 20%. Сделав вычисления для наблюдаемого среднего значения, я получаю $\lambda_1 = 91,26$. Теперь, чтобы понять, что приносит мне больше пользы—увеличение λ или увеличение N я буду сравнивать это значение со случаем, когда $N = 21$, а $\lambda = 100$ и случаем когда $N = 20$, $\lambda = 101$. Первый случай дает значение $\lambda_1 = 91,46$, а второй дает значение $\lambda_1 = 92,21$.

Таким образом, дополнительное наблюдение (большее N) позволило мне обнаружить разницу в средних значениях, которые были меньше на 0,20, в то время как дополнительная частица позволила мне обнаружить разницу в средних значениях, которые были меньше на 0,95. Таким образом, получается, что дополнительная частица делает большее улучшение. Однако в процентном выражении случай с $N = 21$ позволяет мне идентифицировать среднее значение, которое составляет 91,46% от истинного среднего, в то время как случай с $\lambda = 101$ позволяет мне идентифицировать только среднее значение, которое составляет 91,30% от истинного среднего. Теперь оказывается, что дополнительное наблюдение позволяет мне более точно определить разницу в среднем значении этой дополнительной частицы. Так что же лучше? В таблице ниже приведены дополнительные примеры⁸.

	$\lambda=10$	$\lambda=50$	$\lambda=100$	$\lambda=500$
$N=5$	5.1 (51%)	38.2 (76%)	82.9 (83%)	460.8 (92%)
$N=10$	6.4 (64%)	41.5 (83%)	87.8 (88%)	472.1 (94%)
$N=20$	7.4 (74%)	43.9 (88%)	91.3 (91%)	480.2 (96%)
$N=50$	8.3 (83%)	46.1 (92%)	94.4 (94%)	487.4 (98%)

Как мы должны интерпретировать эту таблицу? Похоже, что увеличение N улучшает нашу точность больше, чем увеличение λ , например, удвоение N с 10 до 20 повышает точность на 10% (с 64% до 74%) для наименьшего значения λ . Однако, удвоение λ от 50 до 100 увеличивает точность только на 7% (от 76% до 83%) для наименьшего значения N . По мере того, как N и λ увеличиваются, влияние на точность любого из этих параметров становится менее значительным.

Этот пример еще раз подтверждает точку зрения, которая несколько раз озвучивалась в этой статье—численные методы и техническая строгость не стоят многого, если проблема не понята. Существует ли на самом деле измеримая разница между 91,46% и 91,30% в пересчете на выход кормовой смеси? Если ответ отрицательный, то решение этих сложных математических задач бесполезно с точки зрения улучшения качества производимого продукта. Далее следуют итоги.

9 Резюме

⁸ Записи в ячейках - это значения для λ_1 , которые мы можем идентифицировать с заявленной уверенностью и процентное соотношение λ_1 к λ , которое указано в скобках. Например, $N = 10$, $\lambda = 10$ показывает, что мы можем определить значение $\lambda_1 = 6,4$, которое составляет 64% от истинного среднего

Все изложенное выше согласуется со следующим:

1. Из смеси должно быть взято не менее 10 проб. Результаты улучшатся, если образцы будут взяты из нескольких мест в разное время. Больше образцов всегда лучше, но стоимость отбора проб должна учитываться при определении их количества.
2. Количество добавляемых частиц-индикаторов должно определяться с учетом массы каждого образца таким образом, чтобы ожидаемое количество частиц было не меньше 10. Предположим, что вес образца составляет примерно 100 г, тогда 1 частица или $\frac{1}{2}$ частицы должна приходиться на 1 грамм корма - что соответствует нижней границе. Это привело бы к ожидаемому количеству частиц от 50 до 100 на 100 г образца.
3. Масса образца должна быть небольшой и удобной для забора, при этом не надо упускать из вида обстоятельство, что масса образца влияет на ожидаемое количество частиц. Скорее всего, вес от 50г до 250г представляется целесообразным. Опять же, эта масса должна быть определена в сочетании с количеством добавленных частиц.
4. Образцы должны быть взяты из нескольких мест внутри смесителя. Случайные местоположения допустимы, но случайный выбор места выборки ухудшит способность предсказывать перемещение частиц.