Глава 3. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

СТАТИСТИЧЕСКИЕ ВЫВОДЫ

Основной целью статистического анализа является выяснение свойств изучаемой генеральной совокупности. *Генеральную совокупность* можно рассматривать как полный набор объектов, с которым связана изучаемая проблема. Каждый объект генеральной совокупности называется ее элементом, а соответствующее измерение, произведенное на каждом элементе, – наблюдением. Если генеральная совокупность бесконечна или содержит много элементов, то процедура статистического анализа состоит в том, чтобы отобрать из генеральной совокупности подмножество из N элементов (называемое выборкой объема N), исследовать его свойства, а затем обобщить их на всю генеральную совокупность. Иными словами, задача статистического анализа состоит в получении наилучших в некотором смысле выводов по ограниченному числу наблюдений. Такое обобщение результатов на генеральную совокупность называется *статистическим выводом*.

Статистический вывод можно рассматривать как метод получения утверждений относительно неизвестных параметров изучаемой генеральной совокупности. Эти утверждения можно подразделить на оценивание и проверку гипотез. Оценивание параметров имеет дело с вычислением по выборке точечных и интервальных (доверительных интервалов, включающих истинное значение параметра) оценок. Проверка гипотез заключается в установлении справедливости утверждений (статистических гипотез) относительно параметров генеральной совокупности. Эти два раздела теории статистических выводов – оценивание и проверка гипотез – являются частными случаями общей *задачи принятия решений*.

Рассматривая задачу статистических выводов, подразумевается, что имеется полученная каким-либо образом выборка (*совокупность*). Выборочная совокупность образуется конечным числом наблюдений. Основное требование, предъявляемое к выборке, хорошо представлять генеральную совокупность (быть *репрезентативной*). Обычно это случайная выборка. Формально выборка объема N есть набор реализаций N независимых, одинаково распределенных случайных величин. Простая случайная выборка объема N – это выборка, полученная таким образом, что любая возможная выборка такого объема имела бы одинаковую вероятность быть извлеченной из генеральной совокупности. Таким образом достигается адекватность структуры выборки структуре генеральной совокупности.

Практически не существует стандартного метода получения простой *случайной выборки* из бесконечной генеральной совокупности. Поэтому исследователь вынужден ограничиться конечным подмножеством генеральной совокупности.

Под ошибками репрезентативности, которые могут быть как случайными, так и систематическими, понимают отклонения характеристик выборочной совокупности от соответствующих параметров генеральной совокупности. Эти ошибки, вообще говоря, являются результатом нарушения основного принципа случайной выборки. Например, если палеомагнитная коллекция составлена только из осадочных пород, то, возможно, что наклонение полученных компонент намагниченности будет занижено. Естественно, чтобы контролировать такую возможность, необходимо проводить отбор также и магматических горных пород. Для уменьшения влияния случайных ошибок, например связанных с палеовековыми вариациями, палеомагнитная коллекция должна представлять достаточно много стратиграфических уровней.

При отборе палеомагнитной коллекции могут быть составлены *направленная* или *смешанная* выборки. Отбор направленной выборки осуществляется по ранее установленному принципу, когда он кажется полезным и разумным. Такого рода выборка требует предварительных знаний о генеральной совокупности. Например, после многолетних палеомагнитных исследований рифейских отложений Южного Урала (см., например [Данукалов, Комиссарова, Михайлов, 1982]), было решено для выполнения поставленных задач (сравнительный анализ времени деформаций этого региона) проводить отбор только красноцветных известняков катавской свиты, дающих наиболее однозначные результаты и являющихся ключевым объектом для анализа имеющихся рифейских палеомагнитных определений [Шипунов, 1991, 1995б; Shipunov, 1997]. Смешанная выборка сочетает в себе принципы случайной и направленной выборок. Сущность такого вида выборки состоит в подразделении генеральной совокупности на слои (типичные группы по некоторым заранее установленным признакам) и последующем случайном отборе элементов в выборочную совокупность. При отборе палеомагнитных коллекций чаще всего используют последний вид выборки. Примером могут служить целенаправленный отбор образцов из изменивших свое залегание горных пород (крыльев складок) и случайный отбор из моноклинальных структур с достаточно хорошей представительностью различных стратиграфических уровней.

ТОЧЕЧНОЕ И ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ

Пусть имеется случайная выборка $x_1, ..., x_N$ реализаций случайных величин из генеральной совокупности с плотностью (или законом) распределения вида $f(x, \Theta_1, \ldots, \Theta_k)$, зависящей от k параметров. Например, в случае распределения Фишера для векторов на единичной сфере, функция Φ[μ,κ] зависит от двух параметров – μ (среднего вектора) и κ (кучности).

Требуется оценить один или несколько параметров Θ_i по выборке. Каждая однозначно определенная функция $g(x_1, \ldots, x_N)$, зависящая от наблюдений и не зависящая от неизвестных параметров, которая выбрана для оценки какого-либо параметра распределения, называется значением *точечной оценки*. Для одного и того же параметра совокупности можно построить различные оценки (см. ниже примеры 3.1 – 3.3).

Различают следующие свойства оценок параметров распределений. Оценка неизвестного параметра может быть *несмещенной*; тогда математическое ожидание оценки равно точному значению параметра ($\Theta = \Theta_0$). Так, например, обычно используемая оценка среднего направления совокупности векторов, распределенных в соответствии с законом Фишера, является несмещенной. Напротив, применяемая обычно оценка кучности того же распределения $\kappa = (N–1)/(N–R)$ [Fisher, 1953], как утверждается в [Mardia, 1972; Watson, 1983] – является смещенной. Несмещенной оценкой кучности является $\kappa = (N–2)/(N–R)$ [McFadden, 1980]. Смещенными также являются оценки направления намагниченности в различных вариантах метода пересечения кругов перемагничивания [Schmidt, 1985; Баженов, Шипунов, 1990].

Несмещенные оценки обладают следующими желательными свойствами [Венецкий, Венецкая, 1979]: если имеется серия независимых несмещенных оценок, то их среднее арифметическое также не смещено. С другой стороны, среднее смещенных оценок является, вообще говоря, смещенной оценкой независимо от того, как много оценок осреднялось. Иными словами, требование несмещенности гарантирует отсутствие систематических ошибок при оценке параметров. Это очень важное свойство, которое следует учитывать при анализе палеомагнитных данных. Примером тому может служить среднее направление намагниченности, полученное для осадочных пород, в которых нередко наблюдается занижение наклонения. В этом случае наклонение среднего вектора также будет занижено по сравнению с истинным значением наклонения палеомагнитного поля. Также смещенными могут быть оценки при совместном использовании кругов перемагничивания и конечных точек [McFadden, McElhinny, 1988].

Оценка, кроме того, должна быть состоятельной. *Состоятельность* описывает поведение точечной оценки при неограниченном росте числа наблюдений. Состоятельная оценка при увеличении размера выборки до бесконечности сходится по вероятности к истинному значению параметра, т.е. для сколь угодно малой величины $\varepsilon > 0$ вероятность того, что величина оценки отличается от истинного значения параметра меньше чем на ε, стремится к единице. Поэтому для состоятельной оценки, построенной на основе большого числа наблюдений, вероятность того, что значение оценки будет отличаться от истинного значения, достаточно мала.

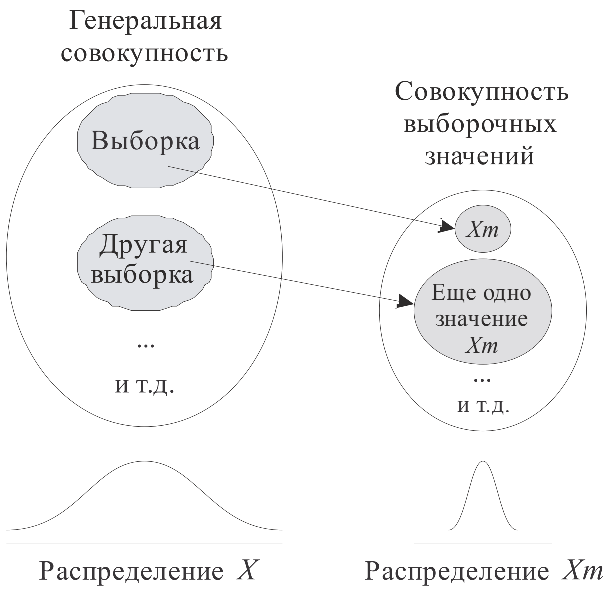

Рис. 3.1. Извлечение выборок из генеральной совокупности

Так как оценка вычислена по случайной выборке, то она сама также является случайной величиной, имеющей некоторый разброс возле истинного значения параметра (т.е. имеет некоторое распределение). Поэтому истинный параметр несколько отличается от вычисленной по результатам наблюдений оценки (рис. 3.1). Следовательно, было бы ошибкой приравнивать истинное значение параметра численному значению оценки.

Как правило, исследователь старается получить такую несмещенную оценку некоторого параметра распределения, дисперсия которой минимальна; такая оценка называется *эффективной*.

При сравнении оценок [Браунли, 1977], включая смещенные, часто ориентируются на их *квадратичный риск*, который определяется как квадрат разности Θ и Θ_0. Легко проверить, что эта величина равна сумме дисперсии оценки и квадрата смещения.

На рис. 3.2 в качестве примера показаны распределения трех конкурирующих оценок Θ_1, Θ_2, Θ_3 для одного параметра Θ. Первые две – имеют преимущество перед Θ_3 в том смысле, что их математическое ожидание равно истинному значению параметра Θ. Оценка Θ_3, напротив, смещена. Оценка Θ_1 предпочтительнее Θ_2 в том смысле, что в малой окрестности Θ плотность оценки Θ_1 больше плотности оценки Θ_2, и, следовательно, она будет чаще принимать значения из этой близкой к истинному значению параметра окрестности. Но, с другой стороны, на хвостах распределений Θ_1 и Θ_2 наблюдается обратная картина, и поэтому большие отклонения от Θ будут также чаще у оценки Θ_1. Если обратить внимание на дисперсию оценок Θ_1, Θ_2, Θ_3 как на меру их разброса, то видно, что наименьшую дисперсию имеет смещенная оценка Θ_3, т.е. она является наилучшей оценкой в этом смысле. Может оказаться, что *квадратичный риск* (сумма дисперсии оценки и квадрата смещения) оценки Θ_3 будет меньше, чем для двух других.

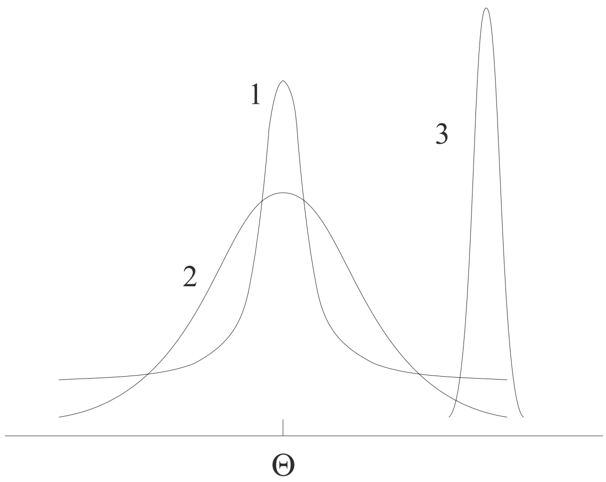

Рис. 3.2. Распределения трех вариантов оценок для одного параметра

После того как получена точечная оценка параметра, желательно получить данные относительно ее надежности. Существует понятие *доверительного интервала*, внутри которого с определенной вероятностью находится истинное значение параметра. Эта вероятность является мерой доверия к тому, что интервал содержит истинное значение параметра. Обычно пользуются *доверительной вероятностью* 95%. В палеомагнитологии вычисляют, например радиус доверительного круга α_{95} для среднего направления совокупности векторов или максимальный и минимальный радиусы овала доверия для оценок положения виртуального геомагнитного полюса и для пересечения кругов перемагничивания. Если говорят, что круг радиуса α_{95} накрывает истинное значение, то это означает, что это утверждение имеет место в среднем в 95% случаев.

Замечание относительно оценки палеомагнитных параметров. Применяя рассмотренные в данном разделе виды оценок и их характеристики в палеомагнитном анализе предполагают, что анализируемая совокупность векторов намагниченности однокомпонентна. В этом случае оценка, например, среднего направления намагниченности характеризует в среднем направление палеомагнитного поля, в котором образовалась эта намагниченность. Напротив, если изучаемая выборка векторов является суперпозицией, например, доскладчатой и послескладчатой компонент намагниченности, то оценка среднего направления фиктивна и независимо от объема выборки не соответствует направлению реального геомагнитного поля.

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Определение среднего направления для совокупности векторов можно свести, к решению задачи методом наименьших квадратов (МНК). Пусть задана каким-либо образом мера отклонения эмпирических данных (векторов выборки) от теоретической модели. Для имеющейся совокупности векторов намагниченности моделью является искомое направление, которое в предположении однокомпонентности намагниченности описывает направление палеомагнитного поля, а мерой отклонения может служить, например, функция от неизвестного направления, отражающая степень различия между исходными векторами и искомым направлением.

Рассмотрим суть метода наименьших квадратов на примере определения среднего направления для совокупности из N векторов [Васильев, 1980]. Все эти векторы, как это принято в палеомагнитологии, единичной длины. Каждый вектор или, что тоже самое точка на единичной сфере, определяется тремя декартовыми координатами x_i, y_i, z_i, где $i = 1, ..., N$.

Пример 3.1. Требуется найти точку (x_0, y_0, z_0), сумма квадратов расстояний (Е на рис. 3.3) от которой до исходных точек минимальна. Иначе говоря, требуется минимизировать функцию

$$S = \sum_{i=1}^{N} \left\{ (x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2 \right\}. \tag{3.1}$$

Здесь в фигурных скобках представлен квадрат евклидова расстояния. В точке экстремума частные производные равны нулю. Поэтому продифференцируем функцию S по неизвестным параметрам и получим систему уравнений

$$\begin{aligned} \partial S / \partial x_0 &= -\sum (x_i - x_0) = 0 \\ \partial S / \partial y_0 &= -\sum (y_i - y_0) = 0 \\ \partial S / \partial z_0 &= -\sum (z_i - z_0) = 0 \end{aligned} . \tag{3.2}$$

решая которую, имеем

$$x_0 = \left(\sum x_i\right)/N, \quad y_0 = \left(\sum y_i\right)/N, \quad z_0 = \left(\sum z_i\right)/N . \tag{3.3}$$

Можно показать [Васильев, 1980], что в этой точке функция S достигает своего глобального минимума в трехмерном евклидовом пространстве.

Заметим, что длина полученного вектора

$$r = \frac{1}{N} \sqrt{\left(\sum x\right)^2 + \left(\sum y\right)^2 + \left(\sum z\right)^2} \le 1 ,$$

причем равенство достигается при совпадении всех исходных векторов (иными словами полученная точка как правило лежит внутри единичной сферы). Любопытно отметить, что минимизируемая величина в точке минимума принимает значение $N - R$ (R – длина суммарного вектора), что является знаменателем в хорошо известной формуле для кучности распределения Фишера; следовательно, оценка кучности – нормированная на объем выборки обратная величина S. Т.е. относительно вектора–решения исходные векторы группируются максимальным образом.

Пример 3.2. В палеомагнитологии часто приходится решать задачу МНК с ограничениями. Рассмотрим предыдущую задачу, но решение будем искать на единичной сфере, т.е. при условии

$$x_0^{\,2} + y_0^{\,2} + z_0^{\,2} = 1 . \tag{3.4}$$

В этом случае задачу можно свести к безусловной минимизации так называемой функции Лагранжа [Ильин, Позняк, 1971]

$$S = \sum \left\{ (x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2 \right\} - \lambda \left(x_0^{\,2} + y_0^{\,2} + z_0^{\,2} - 1 \right). \tag{3.5}$$

Здесь λ – неизвестный множитель Лагранжа. После дифференцирования получим систему

$$\begin{aligned} &\partial S/\partial x_0 = -2\sum(x_i - x_0) - 2\lambda x_0 = 0 \\ &\partial S/\partial y_0 = -2\sum(y_i - y_0) - 2\lambda y_0 = 0. \\ &\partial S/\partial z_0 = -2\sum(z_i - z_0) - 2\lambda z_0 = 0 \end{aligned} \qquad (3.6)$$

решая которую, имеем

$$x_0 = \sum x_i/(N-\lambda),\quad y_0 = \sum y_i/(N-\lambda),\quad z_0 = \sum z_i/(N-\lambda). \qquad (3.7)$$

В соответствии с условием (1.2.4) λ = $N - R$ и, следовательно

$$x_0 = \sum x_i/R,\quad y_0 = \sum y_i/R,\quad z_0 = \sum z_i/R.$$

В рассмотренном примере было использовано необходимое условие минимума функции. Однако, необходимое условие максимума функции такое же. Чтобы определить, с чем в результате мы имеем дело, строго говоря, получив соответствующую оценку, необходимо исследовать вторые производные. Например, подставляя условие нахождения решения на единичной сфере для определения множителя Лагранжа и извлекая квадратный корень, было использовано одно их двух возможных решений. Второе решение

$$x_0 = -\sum x_i/R,\quad y_0 = -\sum y_i/R,\quad z_0 = -\sum z_i/R$$

определяет точку максимума, т.е. наиболее удаленную от исходных данных точку на поверхности единичной сферы.

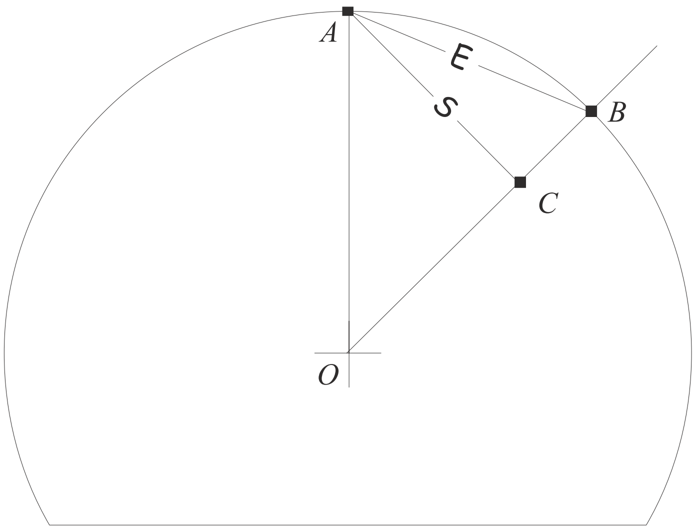

Рис. 3.3. Иллюстрация применения различных метрик для оценки расстояния между двумя векторами *OA* и *OB*. E – евклидово расстояние между точками *A* и *B*, S – метрика, связанная со скалярным произведением (расстояние по перпендикуляру от точки *A* до направления *OB*)

Пример 3.3. В предыдущих примерах в качестве метрики использовалось евклидово расстояние (E на. рис 3.3). В качестве параметра, характеризующего сходимость данных, можно рассмотреть также и другие метрики, например метрику (S на. рис 3.3), связанную со скалярным произведением векторов, например синус угла между искомым направлением и каждым исходным вектором [Девис, 1990]. Такая метрика используется, например, при совместном анализе больших кругов и направлений намагниченности [Kirschvink, 1980].

В этом случае минимизация функции

$$\sum\left\{1-\left(x_0x_i+y_0y_i+z_0z_i\right)^2\right\}=N-\sum\left(x_0x_i+y_0y_i+z_0z_i\right)^2$$

равносильна поиску максимума функции

$$S=\sum\left(x_0x_i+y_0y_i+z_0z_i\right)^2. \tag{3.8}$$

Нулевой вектор является тривиальным решением этой задачи, поэтому в этом случае также приходится искать решение с условием (3.4), т.е. искать максимум функции Лагранжа

$$S=\sum\left(x_0x_i+y_0y_i+z_0z_i\right)^2-\lambda\left(x_0{}^2+y_0{}^2+z_0{}^2-1\right). \tag{3.9}$$

Дифференцирование по неизвестным параметрам сведет задачу к хорошо известной проблеме на собственные значения, возникающей при отыскании решения методом пересечения больших кругов (см., например [Шипунов, 1993]), т.е. к решению системы однородных уравнений

$$\begin{aligned}&x_0\left(\sum x^2-\lambda\right)+y_0\sum xy+z_0\sum xz=0\\&x_0\sum xy+y_0\left(\sum y^2-\lambda\right)+z_0\sum yz=0\\&x_0\sum xz+y_0\sum yz+z_0\left(\sum z^2-\lambda\right)=0\end{aligned}. \tag{3.10}$$

Собственный вектор, соответствующий максимальному собственному значению, будет являться искомым решением (о собственных значениях и собственных векторах см. также Главу 2).

Следует отметить, что оценки направления среднего вектора, полученные в первых двух примерах, совпадают. Третья оценка, связанная со скалярным произведением векторов, несколько отличается от первых двух, но эти различия невелики, особенно при больших значения кучности, т.е. при значениях кучности, обычно используемых при палеомагнитных исследованиях (например, κ > 10 – 20).

МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Для получения оценки неизвестного параметра естественно пытаться найти такое его значение, при котором вероятность реализации имеющейся в наличии выборки будет максимальна.

Пусть $x_1, \ldots, x_N$ – случайные переменные выборочной совокупности, взятой из генеральной совокупности, имеющей плотность распределения $f(x, \Theta)$, вид которой полагается известным. Необходимо найти неизвестный параметр Θ распределения $f(x, \Theta)$ по имеющейся выборке.

Вероятность события, заключающегося в том, что отдельное значение выборки, скажем x_i, попадает в интервал $(x_i–\varepsilon/2, x_i+\varepsilon/2)$, равна приблизительно $f(x_i)\varepsilon$, а вероятность того, что первое значение попадет в первый интервал, второе – во второй и т.д., задается выражением

$$P=f(x_1)\cdot f(x_2)\cdot\ldots\cdot f(x_N)\cdot\varepsilon^N. \tag{3.11}$$

Разделив обе части равенства на ε^N и взяв натуральный логарифм, получим

$$L=\ln\left(P/\varepsilon^N\right)=\sum_{i=1}^{N}\ln\left[f(x_i)\right]. \tag{3.12}$$

Последняя функция известна как *функция правдоподобия* (или логарифм функции правдоподобия). Метод максимального правдоподобия заключается в выборе неизвестного параметра Θ, приводящего к максимальному значению функцию правдоподобия L.

Пример 3.4. Пусть (x_i, y_i, z_i) $i = 1, \ldots, N$ выборка случайных единичных векторов из распределения Фишера Ф$[x_0, y_0, z_0, \kappa]$, имеющего функцию плотности [Mardia, 1972]

$$\left(\frac{\kappa}{4\pi \operatorname{sh}\kappa}\right)\exp\{\kappa(xx_0 + yy_0 + zz_0)\},\quad \kappa > 0. \qquad (3.13)$$

Здесь использована замена cosφ в обычно используемой формуле для плотности распределения Фишера (см. например [Храмов и др., 1982]) на его выражение через скалярное произведение единичных векторов.

Обозначим

$$R_x = \sum x_i,\quad R_y = \sum y_i,\quad R_z = \sum z_i,\quad R = \sqrt{R_x^{\,2} + R_y^{\,2} + R_z^{\,2}}.$$

Логарифм функции правдоподобия для (3.13)

$$\ln L = const + N\ln\kappa - N\ln(\operatorname{sh}\kappa) + \kappa(x_0R_x + y_0R_y + z_0R_z), \qquad (3.14)$$

где

$$x_0^{\,2} + y_0^{\,2} + z_0^{\,2} = 1. \qquad (3.15)$$

Необходимо найти максимум функции правдоподобия (3.14) при условии (3.15). Поэтому напишем как и в примерах 3.2 и 3.3 функцию Лагранжа

$$\text{const} + N\ln\kappa - N\ln(\operatorname{sh}\kappa) + \kappa(x_0R_x + y_0R_y + z_0R_z) - \lambda(x_0^{\,2} + y_0^{\,2} + z_0^{\,2} - 1).$$

Возьмем частные производные по неизвестным параметрам x_0, y_0, z_0, λ и приравняем их нулю

$$\partial \ln L / \partial x_0 = \kappa R_x - 2\lambda x_0 = 0$$

$$\partial \ln L / \partial y_0 = \kappa R_y - 2\lambda y_0 = 0$$

$$\partial \ln L / \partial z_0 = \kappa R_z - 2\lambda z_0 = 0$$

$$\partial \ln L / \partial \lambda = (x_0^{\,2} + y_0^{\,2} + z_0^{\,2} - 1) = 0$$

Отсюда вытекает, что

$$x_0 = \kappa R_x / 2\lambda,\quad y_0 = \kappa R_y / 2\lambda,\quad \mathrm{z}_0 = \kappa R_z / 2\lambda$$

и, учитывая условие (3.15), имеем в результате $x_0 = R_x/R$, $y_0 = R_y/R$, $z_0 = R_z/R$.

Таким образом, оценка генерального среднего направления для распределения Фишера, полученная методом максимального правдоподобия, совпадает с оценкой методом наименьших квадратов с ограничением (пример 3.2). Отметим принципиальное отличие в применении этих двух методов. В методе наименьших квадратов никак не используется вид распределения единичных векторов, тогда как в функцию правдоподобия явным образом входит функция плотности распределения Фишера (3.13). Оценки, полученные методом наименьших квадратом, являются несмещенными. Кроме того, это оценки с наименьшей дисперсией, так как минимизируется сумма квадратов отклонения.

Пример 3.5. Найдем оценку максимального правдоподобия для второго параметра распределения Фишера – кучности κ [Mardia, 1972]. Возьмем частную производную от логарифма функции правдоподобия по этому неизвестному параметру

$$\partial \ln L / \partial \kappa = N/\kappa + \left(x_0 R_x + y_0 R_y + z_0 R_z\right) - N\left[\ln(\operatorname{sh}\kappa)\right]' = 0$$

Заметим, что производная от ln(shκ) есть cthκ, и используя оценку для среднего направления, полученную в предыдущем примере, получим выражение, определяющее кучность

$$r = R/N = \operatorname{cth}\kappa = \mathrm{L}(\kappa). \qquad (3.16)$$

Последняя функция носит название функции Ланжевена (см., например [Watson, 1983]). Для больших r ($r > 0.9$) гиперболический котангенс приблизительно равен единице и оценку максимального правдоподобия можно выразить в явной форме относительно кучности

$$\kappa = N/(N-R) = 1/(1-r). \qquad (3.17)$$

Заметим, что оценки максимального правдоподобия нередко бывают смещенными (аналогично оценке дисперсии одномерного нормального распределения – см., например [Браунли, 1977]). Однако, это смещение можно достаточно легко устранить.

ПРОВЕРКА ГИПОТЕЗ

Статистическая гипотеза – это утверждение относительно значений одного или более параметров данного распределения генеральной совокупности или о самой форме распределения. Под проверкой статистической гипотезы понимается процедура, с помощью которой на основе исходной выборки пытаются установить, следует ли принять это утверждение, называемое нулевой гипотезой H_0, или отвергнуть его. Обычно под нулевой гипотезой рассматривается некоторое выдвигаемое в ходе исследования высказывание (рабочая гипотеза), которым исследователь дорожит и не хочет, чтобы слишком часто оно отвергалось. Справедливость нулевой гипотезы проверяется по имеющейся выборке. Если нулевая гипотеза и имеющиеся данные согласуются с большой степенью правдоподобия, то считается, что H_0 не противоречит им. В противном случае нулевая гипотеза отклоняется.

При выполнении процедуры проверки статистической гипотезы используют некоторую подходящую числовую величину (*статистику критерия*), для которой известна функция распределения вероятностей при условии выполнении нулевой гипотезы H_0. Это распределение, называемое также *нулевым*, унимодально: область, близкая к моде распределения имеет высокую вероятность, тогда как хвосты распределения – это области малых вероятностей. Если величина *критериальной статистики* попадает в область нулевого распределения, имеющую большую вероятность, то можно сделать вывод о том, что выборка не противоречит проверяемой нулевой гипотезе (она согласуется с H_0). Напротив, если наблюдается крайнее, почти невероятное при выполнении H_0 значение величины критерия, то это следует рассматривать как явное расхождение с H_0.

Такая интерпретация результатов тестирования основана на общем принципе, гласящем, что предположение должно быть отвергнуто, если имеется противоречащий пример, но не обязательно должно быть принято, если такого примера найти не удалось. Представленное рассуждение является использованием закона контрапозиции в логике [Справочник..., 1989], в соответствии с которым для произвольных событий или высказываний **A** и **B**, если из **A** вытекает **B**, и не выполняется **B**, то несправедливо и **A**. С другой стороны, предложение “если из **A** следует **B**, и **B** выполняется, то также справедливо и **A**” является ложным, т.к. справедливость **B** может быть следствием не только выполнения **A**, но и в результате наступления другого события (например, **C**).

Пример 3.6. *Тест складки для двух крыльев*. Если намагниченность пород складки доскладчатая (событие **A**), то средние направления векторов намагниченности в стратиграфической (древней) системе координат для крыльев складки *приблизительно* совпадут между собой (событие **B**). Если в результате проведения теста складки направления намагниченности окажутся *сильно* различающимися (событие **B** не выполняется), то из этого следует, что намагниченность нельзя признать доскладчатой (отрицание **A**). Если же в результате проведения теста складки направления намагниченности будут признаны не различающимися, то это еще не означает доскладчатости намагниченности. Например, это может быть связано с малым различием в элементах залегания пластов для крыльев складки. В связи с этим такой результат теста обычно интерпретируется как согласие с выдвинутой гипотезой, т.к. отсутствуют достаточные основания для ее фальсификации. Подробности о тесте складки см. Главу 6.

Пример 3.7. *Тест конгломератов (галек)*. Если намагниченность в гальках образовалась до их переотложения в виде конгломератов (событие **A**), то величина среднего вектора r будет сравнительно мала (событие **B**). Если при проведении теста галек окажется, что величина r превосходит некоторое критическое значение $r^{\#}$, то это ведет к отрицанию события **A**. С другой стороны, если при тестировании выборочное значение статистики $r < r^{\#}$, то это не означает справедливости нулевой гипотезы; такое значение выборочной статистики может быть обусловлено наличием, например, разнополярной компоненты намагниченности с приблизительно равным количеством образцов разной полярности. Подробнее о тесте конгломератов см. Главу 6.

Применительно к статистике принцип контрапозиции можно сформулировать следующим образом: если **B** есть вероятностное следствие **A**, то отрицание **A** будет являться вероятностным следствием отрицания **B**. Если в качестве суждения **A** взять высказывание “ H_0 верна”, а в качестве суждения **B** – “наблюденное значение критериальной статистики, вероятно, близко к моде нулевого распределения”, то нулевая гипотеза, вероятно, не будет верна, если наблюденное значение статистики критерия сильно удалено от моды нулевого распределения.

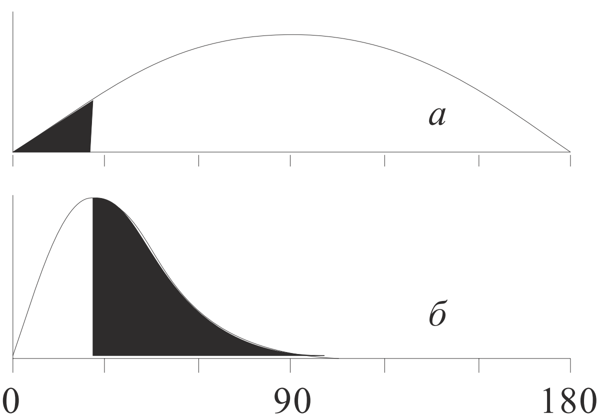

Рис. 3.4. Распределения угла между двумя векторами, извлеченными из равномерного распределения (*а*) и распределения Фишера с кучностью 10 (*б*). Заштрихованы (*а*) – критическая область (область ошибки I-го рода) и (*б*) область ошибки II-го рода

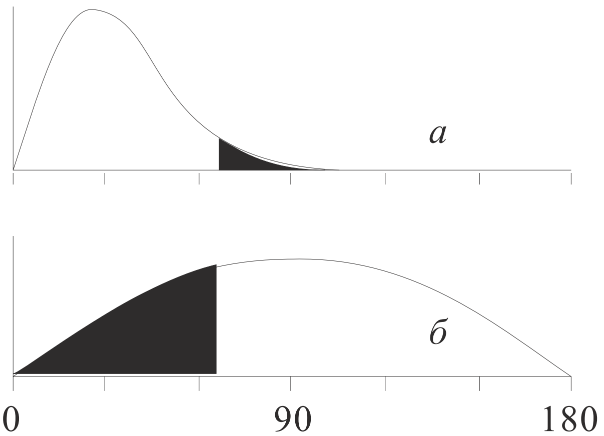

Рис. 3.5. Распределения угла между двумя векторами, извлеченными из распределения Фишера с кучностью 10 (*а*) и равномерного распределения (*б*). Заштрихованы (*а*) – критическая область (область ошибки I-го рода) и (*б*) область ошибки II-го рода

Пример 3.8. Сформулируем, в качестве примера (немного искусственного), задачу проверки гипотезы о параметре распределения (кучности) по выборке, состоящей из двух векторов намагниченности на сфере. Пусть эта выборка получена по двум галькам и мы хотим провести тест галек, т.е. проверить нулевую гипотезу H_0: эти два вектора взяты из равномерного распределения на сфере (или, что то же самое, из распределения Фишера с кучностью $\kappa = 0$). Конкурирующая (альтернативная) гипотеза H_1 пусть будет: $\kappa > 0$. Необходимо сделать выбор между этими гипотезами на основании имеющейся выборки – двух независимо наблюденных векторов намагниченности. Выберем в качестве статистики критерия величину угла φ между этими двумя векторами. На рис. 3.4, *а* приведено нулевое распределение критериальной статистики, т.е. распределение величины φ при условии выполнения H_0. В зависимости от того, в какую область на кривой распределения попадет выборочное значение угла φ, будет принята или отвергнута проверяемая гипотеза H_0. Подмножество возможных значений φ, при попадании в которое будет принята нулевая гипотеза, называется областью принятия нулевой гипотезы. Дополнительная к ней область, при попадании в которую выборочного значения φ нулевая гипотеза отклоняется, называется *критической областью*, а ее граничное значение – *критическим значением*. Естественно, что эта область должна соответствовать малым значениям угла φ.

Критическая область (множество исходов, приводящих к отклонению нулевой гипотезы H_0) выбирается таким образом, чтобы была мала вероятность ошибки α отклонить H_0 в случае, когда она верна (эта ошибка называется ошибкой I-го рода). Эта вероятность α, также называемая уровнем значимости критерия, обычно выбирается равной 0.05 (5%).

Пример 3.8. (*продолжение*). Для рассматриваемого примера критической областью с 5%-ным уровнем значимости будет множество значений $\varphi < \varphi^{\#} \approx 26°$ (заштрихованная область на графике (рис. 3.4, *а*)). Если измеренное значение φ окажется меньше критического, нулевая гипотеза о равномерном распределении отвергается. Вероятность ошибочности этого вывода составит 5%.

Принять нулевую гипотезу H_0, когда она ложна, значит совершить ошибку, которую называют ошибкой II-го рода (вероятность этой ошибки обозначается β). Вероятность дополнительного события, т.е. правильного отклонения H_0, называется мощностью критерия $\pi = 1-\beta$.

Пример 3.8. (*продолжение*). Оценим мощность критерия и вероятность ошибки II-го рода для рассматриваемого примера проверки гипотезы о параметре кучности. Для этого необходимо строго сформулировать альтернативную гипотезу, т.е. знать величину кучности и соответствующее распределение критериальной статистики. Пусть на самом деле $\kappa \geq 10$, а величина вычисленной статистики критерия φ больше критического значения, приблизительно равного 26º. На рис. 3.4, *б* представлено распределение угла φ при $\kappa = 10$, из которого видно, что $\varphi < \varphi^{\#}$ может наблюдаться приблизительно в 50% случаев (заштрихованная область на графике). Мощность критерия для этого случая ($\alpha = 5\%$, $\kappa = 10$) также примерно равна 50%.

Из рассмотренного примера можно сделать следующие выводы. При использование этого критерия: а) в 5% случаев нулевая гипотеза о равномерности распределения будет отклонена ошибочно; б) в 50% случаев нулевая гипотеза будет ошибочно принята если кучность распределения Фишера на самом деле равна 10.

Следует заметить, что увеличения мощности критерия при выбранном уровне значимости α можно добиться увеличением объема рассматриваемой выборки. Так, например, при $N = 10$ и $\kappa = 10$ мощность критерия π близка к 100% (при $N > 2$ в качестве критериальной статистики используется нормализованная длина суммарного вектора $r = R/N$).

Пример 3.9. Рассмотрим те же данные, что и в примере 3.8, но в качестве нулевой гипотезы H_0 возьмем следующее высказывание: два вектора намагниченности взяты из распределения Фишера с кучностью $\kappa = 10$. Альтернативная гипотеза H_1 пусть будет: $\kappa = 0$ (векторы равномерно распределены на сфере). Проверку такой нулевой гипотезы можно производить, например, в случае, когда эти два направления получены по двум образцам из одного пласта и необходимо проверить, имеется ли регулярное направление намагниченности. Заметим, что в этом примере по сравнению с примером 3.8 нулевая и альтернативная гипотезы поменялись местами. Как это скажется на результатах тестирования?

Нулевое распределение для этого случая представлено на рис. 3.5, *а*. Критическое множество значений φ будет располагаться в области больших величин. Критическое значение, при превышении которого нулевая гипотеза будет отклонена, $\varphi^{\#} = 65°$. При отклонении H_0 вероятность ошибочности сделанного вывода составляет 5%. Мощность критерия (при $\kappa = 0$) составляет 72% (см. рис. 3.5, *б*).

Рассмотренные примеры иллюстрируют возможность различной формулировки нулевой гипотезы, в зависимости от того, какой гипотезой исследователь дорожит. Различным отношением исследователя к выдвигаемым при тестировании гипотезам объясняется различие в величинах критических значений φ, используемых для их опровержения.

В [Поллард, 1982] предупреждается об опасности, связанной с применением статистических критериев при анализе одних и тех же данных. Если к одной выборке применить два разных критерия для проверки одной и той же H_0 (или сходных) и в каждом случае использовать уровень значимости, равный, к примеру 5%, то вероятность ошибочного отклонения H_0 хотя бы одним из критериев превосходит 5%. Поэтому следует использовать один их критериев, желательно более мощный. Это напрямую касается теста складки, количество модификаций которого к настоящему времени перевалило за десяток.

Если возникает необходимость проверки двух различающихся гипотез по одной выборке (например, о кучности и среднем направлении совокупности векторов) и для обоих критериев используется 5% уровень значимости, то вероятность ошибочного отклонения хотя бы одной из нулевых гипотез значительно превосходит 5% и может быть близкой к 10%.

В тех случаях, когда неизвестен закон, по которому распределены наблюдения выборки, либо когда значения величин являются ранговыми (в этом случае между значениями величин можно установить лишь соотношения меньше – больше) применяются непараметрические или порядковые критерии. При использования непараметрических критериев рассматриваемые значения сортируют по величине и вычисляют их ранги – т.е. их порядковые номера в отсортированном списке.

Порядковые статистики (критерий Куипера [Мардиа, 1978]) использовались, например, для определения соответствия экспериментальных данных распределению Фишера [Баженов, Рябушкин, 1978]. В методах, описываемых в Главе 6, применяется непараметрическая корреляция Спирмена для установления зависимости между компонентами векторов. Векторы намагниченности не подчиняются нормальному распределению; а в случае выборочных векторов – нормалей к пластам вообще трудно говорить о соответствии какому-либо из известных распределений. Кроме того, предполагаемые зависимости между этими двумя совокупностями векторов в общем случае носят нелинейный характер. Поэтому для выявления таких зависимостей нельзя использовать обыкновенный коэффициент корреляции, предназначенный для исследования линейных связей между нормально распределенными величинами.

Непараметрические методы по мощности, вообще говоря, уступают параметрическим, основанным на известном виде распределения. Преимуществом непараметрических методов является их независимость от вида распределения величин изучаемых выборок (т.е. они могут применяться при различных распределениях) и простота при вычислениях. Если при выполнении непараметрических критериев нулевая гипотеза отвергается или проверка гипотезы показывает соответствие фактических данных нулевой гипотезе, то применение более мощных параметрических методов излишне и нецелесообразно. Если же проверка H_0 не дает определенных выводов, следует (если это возможно) использовать параметрические критерии.

В том случае, когда интересуются как положительными, так и отрицательными отклонениями изучаемых величин, используются двухсторонние критерии. Например, если необходимо установить связь между числовыми характеристиками двух выборок, то вычисляют выборочный коэффициент корреляции r и тестируют нулевую гипотезу H_0: $r = 0$ при альтернативе H_1: $r \neq 0$. Предполагаемая здесь связь может носить как положительный, так и отрицательный характер (в зависимости от знака вычисленного коэффициента корреляции). В этом случае учитывают вероятность попадания вычисленного значения r (при условии выполнения H_0) в оба конца кривой нулевого распределения, которому подчиняется коэффициент корреляции r. Например, если исследователь использует уровень значимости $\alpha = 0.05$, достаточный по его мнению, для проверки значимости корреляции между двумя взаимосвязанными выборками, то верхнее критическое значение $r^{\#}$ будет соответствовать $(1 - \alpha/2) = 0.975$. Нижнее критическое значение $(-r^{\#})$ соответствует $\alpha/2 = 0.025$.

Пример 3.10. Вернемся к рассматриваемому выше примерам 3.8 и 3.9. Пусть попрежнему имеются два вектора и нулевая гипотеза H_0: $\kappa = 0$. Заметим, что слишком малые и слишком большие углы φ между этими двумя векторами маловероятны. Причем углы близкие к 180° могут наблюдаться при присутствии в породе намагниченности обратной полярности. Поэтому имеет смысл рассмотреть в качестве альтернативной гипотезы H_1 наличие регулярной компоненты (возможно обеих полярностей). В этом случае критерий является двухсторонним, а критической областью, при попадании в которую измеренного угла нулевая гипотеза отвергается, являются хвосты нулевого распределения. Тогда нижнее критическое значение $\varphi^{\#}_{\min} = 17°$; вероятность того, что измеренное значение φ будет меньше равна 2.5%. Верхнее критическое значение $\varphi^{\#}_{\max}$ равна 163°; вероятность того, что $\varphi > \varphi^{\#}_{\max}$ также равна 2.5%. Нулевая гипотеза H_0: $\kappa = 0$ отвергается, если φ либо меньше $\varphi^{\#}_{\min}$, либо больше $\varphi^{\#}_{\max}$. Уровень значимости в этом случае 5%.

Рассмотрим типичную процедуру проверки гипотез [Поллард, 1982] на примере проверки гипотезы H_0 о равенстве средних направлений m выборок, взятых из распределений Фишера.

1. Выбирается уровень значимости: $\alpha = 0.05$.

2. Описывается статистическая модель: распределения m выборок векторов соответствуют распределению Фишера с одинаковой кучностью. Это означает, что перед проверкой гипотезы о равенстве средних, необходимо провести проверку гипотезы о соответствии каждой из m выборок распределению Фишера (см., например [Храмов и др., 1982]) и проверку гипотезы о гомогенности этих выборок (равенства их кучностей) [Mardia, 1972]. В случае удовлетворительных результатов этих двух тестов для m выборок выборочные распределения будут соответствовать принимаемой статистической модели.

3. Формулируются нулевая H_0: средние направления m выборок векторов равны и альтернативная гипотезы H_1: средние направления различаются.

4. Выбирается критериальная статистика, распределение которой известно (т.е. численная величина, вычисляемая по некоторой формуле, распределение которой известно в случае выполнения H_o). Критериальная статистика для данного критерия [McFadden, Jones, 1981]

$$F_m = \frac{N-m}{m-1} \cdot \frac{\sum R_i - R^2 / \sum R_i}{2\left(N - \sum R_i\right)}$$

распределена в соответствии с распределением отношения дисперсий (F-распределением) $F[\nu_1, \nu_2, \alpha]$, где N – общее количество векторов в m выборках, R_i – длина суммарного вектора i-той совокупности, R – длина суммарного вектора объединенной выборки, $\nu_1 = 2(m - 1)$ и $\nu_2 = 2(N - m)$ – степени свободы F-распределения.

5. Определяется критическая область по таблицам процентных точек (критических значений) F-распределения с числом степеней $\nu_1 = 2(m–1)$ и $\nu_2 = 2(N–m)$ и уровнем значимости α.

6. Вычисляется значение статистического критерия по наблюдаемым выборкам и делается вывод. Если полученное значение критерия лежит в критической области (т.е. $F_m > F^{\#}$), то следует отклонить нулевую гипотезу и принять альтернативную. В противном случае – принять нулевую гипотезу. α равна вероятности отвергнуть H_0 при условии, что H_0 верна. β равна вероятности принять H_0 при условии, что H_0 ложна. В таблице 3.1 иллюстрируются возможные выводы и вероятности ошибок I-го и II-го рода для этого примера.

Таблица 3.1. Интерпретация результатов и возможные ошибки при проверке гипотезы о равенстве средних

Принятое решение	В действительности	
	Средние равны	Средние не равны
Средние равны	Правильное решение Вероятность $1 - \alpha$	Ошибка II-го рода Вероятность ошибки β
Средние не равны	Ошибка I-го рода Вероятность ошибки α	Правильное решение Вероятность $\pi = 1 - \beta$

Опыт обычно не доказывает того, что та или иная нулевая гипотеза H_0 справедлива – возможно лишь доказать ее неправильность. Чаще всего дается уклончивый ответ на вопрос о том, совместима ли H_0 с наблюдениями или нет, но если результат вычислений величины критерия имеет очень малую вероятность и попадает в критическую область, то это является более или менее определенным указанием на то, что H_0 необходимо отвергнуть. Неотрицательный ответ по H_0 не означает, что она лучшая или единственная. Она представляет собой одно из правдоподобных утверждений, а факты могут лишь опровергнуть выдвинутую H_0, но не могут ее подтвердить. Использование нулевой гипотезы напоминает [Венецкий, Венецкая, 1979] принцип презумпции невиновности в юриспруденции, в соответствии с которым “обвиняемый считается невиновным до тех пор, пока его вина не доказана”.

Обычно применяемые критерии с уровнем значимости α контролируют лишь ошибки первого рода и не определяет степень риска, связанного с ошибками второго рода. Для определения вероятности β ошибки второго рода необходимо сформулировать альтернативную гипотезу таким образом, чтобы можно было определить вероятность принятия H_0 в том случае, если верна H_1, т.е. полностью определить функцию распределение вероятности при выполнении H_1, что не всегда возможно. Вероятность β может оказаться довольно большой (см. примеры 3.8 и 3.9).

Тем не менее, если вычисленная критериальная статистика попала в область принятия гипотезы H_0, то нулевую гипотезу принимают и верят в нее до тех пор, пока последующие наблюдения не заставят изменить мнение. Таким образом, если наблюдения не противоречат H_0, то она остается открытой, так как результаты наблюдений иногда столь же убедительно можно согласовать с множеством других, не выдвинутых гипотез.

Такая процедура проверки гипотез является стандартной и не свободна от недостатков, связанных с несимметричностью влияния ошибок I-го и II-го рода. Часто более удобно производить другую процедуру проверки гипотез, являющуюся в некотором смысле обратной к стандартной. Вместо того, чтобы работать с фиксированным уровнем значимости α и только соглашаться с отклонением или принятием H_0, можно, пользуясь вычисленной критериальной статистикой, определить вероятность α, которому она соответствует. Число α, полученное таким образом, называется P-значением. При такой процедуре проверки гипотезы, отвергая (принимая) гипотезу H_0, известен точный уровень значимости $\alpha = P$, на котором происходит отклонение (принятие) нулевой гипотезы. Такое объяснение результатов тестирования предлагается, например в статье [Баженов, Рябушкин, 1978] при проверке соответствия совокупности векторов распределению Фишера.

Мощность критерия ($\pi = 1 - \beta$) при выбранном уровне значимости зависит от объема выборки N, и, увеличивая его, можно добиться увеличения мощности критерия и уменьшения вероятности ошибочного принятия нулевой гипотезы β.

Рассмотрим влияние объемов выборок на величину мощности критерия [Браунли, 1977] на примере проверки равенства кучностей двух совокупностей векторов. Нулевая гипотеза H_0: $\kappa_1 = \kappa_2$ ($\alpha = 0.05$). Пусть вычисленные значения $\kappa_1 > \kappa_2$. Критериальная статистика κ_1/κ_2 распределена в соответствии с F-распределением $F[\nu_1, \nu_2, \alpha]$, где степени свободы $\nu_1 = 2(N_1 - 1)$ и $\nu_2 = 2(N_2 - 1)$. Пользуясь [Watson, 1956], для альтернативной гипотезы $\kappa_1/\kappa_2 = \varphi$ мощность критерия $\pi(\varphi)$ равна вероятности

$$P\{F[\nu_1, \nu_2, \alpha] > (1/\varphi)\, F[\nu_1, \nu_2, 1-\alpha]\}. \qquad (3.18)$$

Например, если $N_1 = 16$ и $N_2 = 6$, то $F[30, 10, 0.95] = 2.7$, и если альтернативная гипотеза $\varphi = \kappa_1/\kappa_2 = 5$, то мощность критерия

$$\pi = P\{F[30, 10] > 2.7/5\} = P\{F[10, 30] < 5/2.7\} = 0.90.$$

Здесь используется следующее свойство процентных точек F-распределения $1/F[\nu_1, \nu_2, \alpha] = F[\nu_1, \nu_2, 1-\alpha]$. Если же альтернативная гипотеза $\varphi = \kappa_1/\kappa_2 = 1.5$, то мощность критерия

$$\pi = P\{F[30, 10] > 2.7/1.5\} = P\{F[10, 30] < 1.5/2.7\} = 1 - 0.90 = 0.10.$$

Таким образом, можно получить представление о том, как влияет выбор величины отношения кучностей φ в формулировке альтернативной гипотезы на величину мощности критерия.

Для того, чтобы определить необходимые объемы выборок, для которых критерий будет иметь заранее заданную мощность $\pi = 1 - \beta$, приравняем правую часть формулы (3.18) $(1 - \beta)$. Но так как

$$P\{F[\nu_1, \nu_2] > F[\nu_1, \nu_2, \beta]\} = 1-\beta$$

то получается уравнение относительно ν_1 и ν_2

$$\varphi = F[\nu_1, \nu_2, 1-\alpha]/F[\nu_1, \nu_2, \beta] = F[\nu_1, \nu_2, 1-\alpha]\, F[\nu_2, \nu_1, 1-\beta],$$

решая которое с помощью таблиц процентных точек F-распределения получим следующее. Если $\alpha = 0.05$ и $\pi = 1 - \beta = 0.90$, то $F[80, 80, 0{,}95] = 1.45$ и $F[80, 80, 0.90] = 1.33$, а их произведение 1.93. Поэтому для $\varphi = 2$ необходимо взять объемы выборок (при равенстве объемов) $N_1 = N_2 = 39$.

ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ

Численное моделирование или метод Монте–Карло представляет собой способ решения разнообразных задач путем использования случайных величин. Достоинство метода Монте–Карло – простота вычислительных алгоритмов, моделирующих реальные процессы при решении различного рода задач. Эффективное использование численного моделирования возможно только на компьютере; при этом обычно требуется большое число реализаций случайных величин.

При анализе палеомагнитных данных численное моделирование может применяться в трех областях: при анализе некоторой проблемы, получении точечных и интервальных оценок и проверке гипотез.

В программном обеспечении большинства компьютеров имеются так называемые датчики случайных чисел – подпрограммы, которые вычисляют последовательность псевдослучайных чисел, равномерно распределенных на интервале (0, 1). Способы образования последовательностей случайных чисел с заданным законом распределения состоят в преобразовании равномерно распределенных чисел в необходимую последовательность.

Общий способ получения случайных чисел, функция распределения которых $F(x)$, заключается в следующем. Пусть γ – случайное число, равномерно распределенное на интервале (0, 1). Тогда, пользуясь обратной функцией, вычисляют $\delta = F^{-1}(\gamma)$. Величина δ будет распределена в соответствии с требуемым законом. Проблема сводится тогда к достаточно быстрому и точному вычислению функции $F^{-1}(x)$. В тех случаях, когда функция $F^{-1}(x)$ не представима аналитически, приходится для каждого значение γ численно решать некоторое уравнение.

Рассмотрим простой пример получения случайного числа, распределенного равномерно на интервале (a, b). Функция распределения вероятностей имеет вид

$$\mathrm{F}(x) = \begin{cases} 0, & \text{для } x < a \\ (x\text{-}a)/(b-a), & \text{для } a < x < b \\ 1, & \text{для } x > b \end{cases}$$

Обратная функция $\mathrm{F}^{-1} = a + (b-a)x$. Поэтому, если γ – случайное число, равномерно распределенное на (0, 1), то

$$\delta = a + (b-a)\gamma$$

равномерно распределена на (a, b).

Рассмотрим пример получения случайного вектора, распределенного равномерно в некотором круге сферы с центром (D_0, I_0) и радиусом φ_0. Функции распределения вероятностей азимутального и радиального углов, имеют вид:

$$F(\psi) = \psi/2\pi, \quad F(\varphi) = (1-\cos\varphi)/(1-\cos\varphi_0).$$

Соответственно, обратные функции выражаются формулами:

$$\psi = 2\pi \cdot x, \quad \varphi = \arccos[1 - y(1-\cos\varphi_0)].$$

Поэтому, если x и y – два случайных числа, равномерно распределенных на (0, 1), то (ψ, φ) – вектор, равномерно распределенный на сфере в круге радиусом φ_0. Так как пределы изменения радиального угла от 0 до π, то чтобы получить искомое распределение осталось преобразовать φ к виду, используемому при задании палеомагнитных координат векторов: $\varphi^* = \pi/2 - \varphi$, и затем произвести вращение (ψ, φ^*) к центру заданного распределения (D_0, I_0) (см. Главу 1).

Моделирование единичных векторов на сфере из равномерного и фишеровского распределений производят аналогичным образом, используя обратные функции к формулам (2.5 – 2.7) Главы 2.

Пример 3.11. Используем моделирование методом Монте–Карло для определения разброса выборочных оценок, например, определения α_{95} для среднего направления выборки векторов из фишеровского распределения [Fisher, Lewis, Embleton, 1987]. Пусть имеются выборочные оценки D^*, I^*, κ^* для распределения Фишера $\Phi[D_o, I_o, \kappa, N]$. Моделируя M раз распределение $\Phi[D^*, I^*, \kappa^*, N]$ по выборочным значениям D^*, I^*, κ^* и N получим M оценок среднего $D_i, I_i, i = 1, \ldots, M$. Величина M должна быть достаточно большая, например $M = 1000$. Круговая область вокруг центра всех выборок, в которой располагается 95% полученных таким образом оценок среднего направления, есть оценка α_{95}, определенная с помощью численного моделирования.

ПРОВЕРКА ГИПОТЕЗ МЕТОДОМ БАРНАРДА

Этот метод [Barnard, 1963; Marriot, 1979; Мэйндоналд, 1988] используется для нахождения процентных точек распределений (например, для определения критических значений), величины которых трудно или невозможно получить аналитически. Предположим, что необходимо проверить некоторую выборочную статистику z^* на значимость на α%-ном уровне (пусть для определенности большие значения величины z ведут к отклонению нулевой гипотезы H_0), а таблица критических значений z не доступна. Проведя $M - 1$ (M достаточно большое число, например, 1000) испытание по методу Монте–Карло при условии выполнения H_0, получим значения статистики критерия

$$z_1, z_2, \ldots, z_{M-1}.$$

Добавив к ним z^* и упорядочив все эти величины по возрастанию, имеем в результате

$$z_{(1)} < z_{(2)} < \ldots < z_{(M)}.$$

Для определенности в случае равенства поставим z^* на первое место. Тогда согласно H_0 вероятность того, что z^* попадет в число m наибольших величин в отсортированной последовательности равна, m/M. Вероятность ошибки I-го рода для критерия, отклоняющего нулевую гипотезу H_0, если z^* находится среди m наибольших величин, будет, таким образом, $\alpha = m/M$.

Аналогичная процедура может быть применена для вычисления критических значений. При этом проводим M испытаний по методу Монте–Карло при условии выполнения гипотезы H_0 и получаем значения статистики критерия

$$z_1, z_2, \ldots, z_M.$$

Упорядочив все эти величины по возрастанию, имеем в результате

$$z_{(1)} < z_{(2)} < \ldots < z_{(M)}.$$

Тогда критическим значением статистики с уровнем значимости α является $(1 - \alpha)$ M величина в отсортированном списке (здесь как и выше для определенности предполагается, что большие значения величины z ведут к отклонению нулевой гипотезы).

КРИТЕРИИ РАНДОМИЗАЦИИ

Рандоминизированные критерии – это непараметрические процедуры, основанные на истинных наблюдениях и их статистической природе. Они полезны для малых выборок, но могут стать очень громоздкими при увеличении объемов выборок. Разновидностями подобного рода критериев являются “*permutation*” (посредством перемен), использующееся в англоязычной статистической литературе (см., например [Fisher, Lewis, Embleton, 1987]) и “*bootstrap*”, нашедшее применение в палеомагнитной литературе (см., например [Tauxe, Watson, 1994]).

Рассмотрим два примера применения критерия рандоминизации.

Пример 3.12. Опишем кратко процедуру проверки гипотезы о том, что две имеющиеся выборки векторов взяты из одного распределения (не обязательно фишеровского). Исходными данными служат D_{j1}, I_{j1}, n_1, D_{j2}, I_{j2}, n_2, где D_j, I_j, – склонения и наклонения векторов выборок, а n_1 и n_2 – их объемы.

Допустим, что вид и параметры распределения полностью известны. Тогда можно численно смоделировать две выборки объемами n_1 и n_2 из этого распределения. Каждый вектор, например, первой выборки с таким же успехом может оказаться в другой выборке. Иными словами, с равной вероятностью любой из единичных векторов может принадлежать как первой, так и второй выборке. Поэтому не будет ошибкой поменять местами любые пары векторов в выборках. При этом мы никак не учитывали вид и параметры распределения, а использовали только тот факт, что каждая выборка (или каждый вектор в одной из двух выборок) взяты из одной генеральной совокупности.

Поэтому при истинности проверяемой нулевой гипотезе о том, что две имеющиеся выборки векторов взяты из одной генеральной совокупности, можно проделать те же операции по перестановке векторов. Тогда процедура тестирования выглядит следующим образом.

1. Вычисляется угловое расстояние между средними векторами выборок φ_0, которое будет служить статистикой критерия.

2. Много раз (например, $m = 1000$) из объединенной выборки объема $N = n_1 + n_2$ случайным образом составляются две выборки объемами n_1 и n_2 и определяется угловое расстояние между ними φ_j, $j = 1, \dots, m$ (процедура, создающая две случайные выборки из объединенной выборки, должна удовлетворять требованию несмещенности, т.е. каждая полученная реализация должна встречаться в процессе тестирования только один раз [Гудман, Хидетниеми, 1981]).

3. Полученный массив углов φ_j сортируется в порядке возрастания.

4. Производится сравнение исходной величины угла φ_0 с нулевым распределением этого угла, т.е. возможными его значениями при условии выполнения нулевой гипотезы. Если величина φ_0 в некотором смысле соответствует нулевому распределению, то нулевая гипотеза принимается. В противном случае, например, при достаточно большом ее значении (большем некоторого критического), нулевая гипотеза должна быть отвергнута. Критическим значением на 5% уровне значимости служит $0.95m$-ое значение угла в отсортированном списке φ_j.

Пример 3.13. Процедура проверки гипотезы о независимости (значимости коэффициента корреляции) двух связанных одномерных выборок. Пусть имеются две выборки X, Y: (x_1, y_1), (x_2, y_2), … , (x_N, y_N). Основной довод за использование критерия рандомизации состоит в том, что при выполнении нулевой гипотезы о независимости имеющихся выборок каждая пара (x_i, y_i) является просто одной из многих равновероятных пар. Например, другая возможная выборка (x_1, y_2), (x_2, y_3), … , (x_N, y_1) имеет такую же вероятность появления при условии выполнения нулевой гипотезы. Общее число таких возможных различных выборок соответствует числу различных перестановок последовательности Y при неизменной последовательности X и равно $N!$.

Для выполнения теста рандомизации необходимо вычислить некоторую статистику, определяющую корреляцию. Обычно используемая статистика (коэффициент корреляции)

$$r = \frac{\sum(x - x^*)(y - y^*)}{\left[\sum(x - x^*)\sum(y - y^*)\right]^{1/2}}$$

где $x^* = \Sigma x_i/N$, $y^* = \Sigma y_i/N$. Сначала вычислим коэффициент корреляции r для имеющейся выборки; обозначим его r^*. Произведем перестановку в последовательности Y и вычислим выборочную корреляцию для такой новой выборки. Повторим последнюю операцию для всех возможных перестановок и получим $N!$ коэффициентов корреляции $r_1, r_2, \dots, r_M$ ($M = N!$).

Следующий шаг состоит в определении, соответствует ли выборочное значение r^* другим $M-1$ значениям коэффициента корреляции. Для этого отсортируем последовательность r_i в возрастающем порядке и получим

$$r_{(1)} < r_{(2)} < \ldots < r_{(M)}.$$

Предположим для определенности, что нас интересует альтернативная гипотеза о наличии положительной корреляции, т.е. односторонний тест. В этом случае нам следует отклонить гипотезу о независимости исходных выборок если величина r^* значимо больше других сравниваемых значений r; например, при тестировании на α уровне значимости нулевая гипотеза будет отклонена, если r^* больше $(1-\alpha)$ M-го значения в отсортированном списке. Такой результат можно проинтерпретировать как невыполнение основного предположения о том, что все $N!$ возможные выборки равновероятны.

Вычислительные аспекты критериев рандомизации имеют важное значение. Для выборок большого объема трудно вычислить все $N!$ возможные перестановки исходных данных. Эту проблему можно удовлетворительным образом обойти, генерируя достаточно большое число M (скажем, 1000 или 2000) случайных перестановок и вычисляя затем коэффициенты корреляции для соответствующих M выборок. Алгоритм получения таких случайных несмещенных выборок из исходной выборки можно найти, например в [Гудман, Хидетниеми, 1981].

НОМЕНКЛАТУРА ДЛЯ ТАБЛИЧНЫХ ЗНАЧЕНИЙ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ

В статистической практике часто необходимы таблицы функций различных вероятностных распределений. Они нужны, например при тестировании выдвигаемых в процессе исследования статистических гипотез. Критериальные статистики, используемые для этого, распределены, как правило, в соответствии с широко известными распределениями теории вероятностей. Поэтому критические области, при попадании в которую выборочного значения статистики нулевая гипотеза отвергается, принадлежат нижним и верхним хвостам этих распределений. Почти в каждой книге по статистике приводятся таблицы процентилей (процентных точек), которые удовлетворяют следующим выражениям:

$$\mathrm{P}(z > z_\alpha) = \alpha, \quad \mathrm{P}(z < \zeta_\beta) = \beta.$$

Здесь z_α и ζ_β соответственно верхняя и нижняя процентные точки; причем $\zeta_\beta = z_{1-\beta}$.

Однако, использовать таблицы процентилей при вычислениях на компьютерах, не совсем удобно – для этого надо, во-первых, вводить в программы длинные массивы табличных данных, во-вторых, не всегда можно найти табличные значения процентилей для различного уровня значимости.

ЛИТЕРАТУРА

В данной главе использованы следующие работы автора:

Баженов М.Л., Шипунов С.В. (1990) Метод пересечения дуг большого круга: анализ и приложения в палеомагнетизме и тектонике плит // Изв. АН СССР. Физика Земли. 1990. № 1. С.

Шипунов С.В. (1991) О палеомагнетизме катавской свиты Южного Урала // Изв. АН СССР. Физика Земли. 1991. № 3. С. 97-109.

Шипунов С.В. (1993) Основы палеомагнитного анализа: Теория и практика. М.: Наука. 1993. 159 с. (Труды ГИН. Вып. 487)

Шипунов С.В. (1995) Синскладчатая намагниченность: Оценка направления и геологическое приложение // Физика Земли. 1995. № 11. С. 40-47.

Shipunov S.V. (1997) Synfolding magnetization: Detection, testing and geological applications // Geophys. J. Int. 1997. Vol. 130. P. 405-410.

Использованная литература:

Баженов М.Л., Рябушкин П.К. (1978) Применение статистических критериев согласия в палеомагнитных исследованиях // Изв. АН СССР. Физика Земли. 1978. № 7. С. 100-104.

Браунли К.А. (1977) Статистическая теория и методология в науке и технике. М.: Наука. 1977. 408 с.

Васильев Ф.П. (1980) Численные методы решения экстремальных задач. М.: Наука. 1980. 520 с.

Венецкий И.Г., Венецкая В.И. (1979) Основные математико-статистические понятия и формулы в экономическом анализе. М.: Статистика. 1979. 448 с.

Гудман С., Хидетниеми С. (1981) Введение в разработку и анализ алгоритмов. М.: Мир. 1981. 368 с.

Данукалов Н.Ф., Комиссарова Р.А., Михайлов П.Н. (1982) Палеомагнетизм ривея и венда Южного Урала // Стратотип рифея: Палеонтология. Палеомагнетизм. М.: Наука, 1982. С. 121–161. (Тр. ГИН АН СССР; вып. 368).

Девис Дж. (1990) Статистический анализ данных в геологии. М.: Недра. 1990. Кн. 1. 320 с. Кн. 2. 428 с.

Ильин В.А., Позняк Э.Г. (1971) Основы математического анализа. М.: Наука. 1971. Часть I. 600 с.

Мардиа К. (1978) Статистический анализ угловых наблюдений. М.: Наука. 1978. 240 с.

Мэйндоналд Дж. (1988) Вычислительные алгоритмы в прикладной статистике. М.: Финансы и статистика. 1988. 352 с.

Храмов А.Н., Гончаров Г.И., Комиссарова Р.А., Писаревский С.А., Погарская И.А., Ржевский Ю.С., Родионов В.П., Слауцитайс И.П. (1982) Палеомагнитология. Л.: Недра. 1982. 312 с.

Поллард Дж. (1982) Справочник по вычислительным методам статистики. М.: Финансы и статистика. 1982. 346 с.

Справочник по прикладной статистике / Ред. Э. Ллойд, У. Ледерман. М.: Финансы и статистика. Том 1. 1989. 510 с. Том 2. 1990. 526 с.

Barnard G.A. (1963) In discussion following a paper by Bartlett // Journal of the Royal Statistical Society. 1963. B 25. P. 294.

Fisher R.A. (1953) Dispersion on a sphere // Proc. R. Soc. London. 1953. A 217. P. 295-305.

Fisher N.I., Lewis T., Embleton B.J.J. (1987) Statistical analysis of spherical data. Cambridge: Cambridge University Press. 1987. 330 p.

Kirschvink J.L. (1980) The least-squares line and plane and analysis of palaeomagnetic data // Geophys. J. R. astr. Soc. 1980. Vol. 62. P. 699-718.

Mardia K.V. (1972) Statistics of directional data. London: Academic Press. 1972. 357 p.

McFadden P.L. (1980) The best estimate of Fisher's precision parameter // Geophys. J. R. astr. Soc. 1980. Vol. 60. P. 397-407.

McFadden P.L., Jones D.L. (1981) The fold test in palaeomagnetism // Geophys. J. R. astr. Soc. 1981. Vol. 67. P. 53-58.

McFadden P.L., McElhinny M.W. (1988) The combined analysis of remagnetization circles and direct observations in palaeomagnetism // Earth and Planetary Science Letters. 1988. Vol. 87. P. 161-172.

Marriott F.H.C. (1979) Barnard's Monte Carlo tests: How many simulations? // Applied Statistics. 1979. Vol. 28. P. 75-77.

Schmidt P.W. (1985) Bias in converging great circle methods // Earth and Planetary Science Letters. 1985. Vol. 72. P. 427-432.

Tauxe L., Watson G.S. (1994) The fold test: an eigen analysis approach // Earth and Planetary Science Letters. 1994. Vol. 122. P. 331-341.

Watson G.S. (1956) Analysis of dispersion on a sphere // Monthly Notices R. astr. Soc., Geophys. Suppl. 1956. Vol. 7. P. 153-159.

Watson G.S. (1983) Statistics on spheres. University of Arkansas Lecture Notes in the Mathematical Sciences. Vol. 6. New York. John Wiley. 1983. 238 p.