В прошлый раз мы считали, что имеется выборка по распределению, про которое мы уверены, какой вид оно имеет и хотим только найти несколько параметров. В этот раз мы рассмотрим ситуацию, когда у нас есть одна основная гипотеза, полностью задающая это распределение, и мы хотим проверить, насколько она согласуется с наблюдаемыми фактами.

Постановка задачи выглядит так: у нас имеется основная гипотеза (нулевая гипотеза), задающая конкретное распределение. Возможно, что у нас также есть ещё одна гипотеза (альтернативная гипотеза), которая задаёт ещё одно конкретное распределение и является нашим следующим вариантом после нулевой гипотезы. У нас есть выборка, полученная повторными независимыми испытаниями с одним и тем же распределением и мы хотим проверить, что она получена испытаниями с распределением, заданным нулевой гипотезы. При этом возможно два вида ошибок: мы можем отвергнуть нулевую гипотезу, хотя она и была верна (это ошибка первого рода); или же мы можем не отвергнуть нулевую гипотезу, хотя она и неверна (это ошибка второго рода). Разумеется, мы можем исключить один из видов ошибки полностью, если примем решение без учёта выборки. Но хочется добиться того, чтобы риски совершить ошибку первого и второго рода были сбалансированы.

Обычно фиксируется некоторое конкретное ε и задаётся требование, чтобы во-первых, вероятность ошибки первого рода не превышала ε; а во-вторых, вероятность ошибки второго рода при каждом конкретном альтернативном распределении была по возможности меньше. Большая чёткость формулировки требований к ошибке первого рода связана и с тем, что мы склонны сохранить нулевую гипотезу при отсутствии убедительных аргументов, и с тем, что только ошибку первого рода можно оценить точно не задавая альтернативной гипотезы.

Многие критерии можно построить, применив какую-то функцию к выборке и оценив распределение результата (в предположении нулевой гипотезы). После этого мы отвергаем нулевую гипотезу, если выборка попала во множество наименее вероятных выборок с точки зрения значения этой функции (отсечение наименее вероятных выборок производится так, чтобы суммарная вероятность отсечённых составила ε).

Начнём с критерия Колмогорова. Этот критерий позволяет проверить нулевую гипотезу для произвольного распределения без фиксации альтернативной. Определим сначала выборочную функция распределения. Это просто функция распределения, задающая равномерное распределение на выборке; другими словами, если $x_1, \ldots, x_N$ --- выборка, то $F_N(z)$ --- это доля элементов выборки, лежащих слева от z. Можно спросить, как связана выборочная функция распределения с истинной функцией распределения. Оказывается, что точная верхняя грань разности их значений по $\mathbb{R}$ с ростом N стремится к 0 по вероятности. Точнее говоря, определим $D_N = sup|F_N(x) - \Phi(x)|$, где Φ --- истинная функция распределения. Тогда теорема Колмогорова утверждает, что $P(\sqrt{N}D_N < z) \to K(z)$ с ростом N, где $K(z) = 1 + 2\sum\limits_{k\geqslant 1}(-1)^k e^{-2k^2z^2}$. Более того, последовательность $P(\sqrt{N}D_N < z)$ не зависит от исходного распределения Φ. Разумеется, критерий Колмогорова предписывает отбрасывать нулевую гипотезу когда $\sqrt{N}D_N > K^{-1}(1-\varepsilon)$ (при наличии достаточных вычислительных ресурсов можно даже учесть конкретное N в выборе порогового значения).

Доказательство теоремы Колмогорова состоит из двух частей: доказательства, что распределение D_N не зависит от Φ и вычисления $K(z)$ для некоторого конкретного распределения Φ, наиболее упрощающего вычисления. Вторую часть доказательства мы опустим, чтобы не вдаваться в детали техники, относящейся к математическому анализу.

Доказать независимость D_N от Φ можно, рассмотрев его поведение при замене переменных. Так как Φ и F_N

формулируются в терминах неравенств на значения случайных величин, замена переменных обязана быть монотонной. С другой стороны, никаких других требований нет: функция распределения растянется или сожмётся на каких-то участках, но вероятность попасть между образами двух точек останется равной вероятности попасть между исходными точками. А правильной заменой переменных можно привести Φ к фиксированному виду.

Другой способ сформулировать это рассуждение выглядит так: отразим график Φ относительно прямой $y = x$, выберем на отрезке $[0; 1]$ случайную точку по равномерному распределению и найдём её образ относительно отражённой Φ. Ясно, что так мы получим точку, распределённую по Φ. С другой стороны, D_N можно посчитать как максимум из длин отрезков, на которые N случайно выбранных точек разобьют отрезок $[0; 1]$ (после применения обращённой Φ получится ровно исходное определение).

Для случая, когда у нас есть две гипотезы и мы хотим их сравнить, можно применить критерий Неймана-Пирсена. Рассмотрим его в случае, когда у нас нулевая и альтернативная гипотезы задают распределения на вещественных числах с плотностью. Рассмотрим неравенство $f_0(x) - \lambda f_1(x) \geqslant 0$, где f_0 и f_1 --- плотности вероятности выборок, задаваемые нулевой и альтернативной гипотезами. Найдём такое λ, что вероятность невыполнения этого неравенства в предположении нулевой гипотезы станет равна ε. Будем отвергать нулевую гипотезу в случае нарушения неравенства.

Покажем, что это оптимальный с точки зрения ошибки второго рода критерий для сравнения гипотез при заданной вероятности ошибки первого рода. Действительно, всё множество возможных выборок мы разбили на два подмножества --- подмножество сохранения нулевой гипотезы и подмножество её отвержения. Если мы изменим разбиение, мы заменим некоторое подмножество первой области на подмножество второй области с той же вероятностью в предположении нулевой гипотезы. Назовём эту вероятность δ. Но заметим, что у выкинутого подмножества вероятность в предположении альтернативной гипотезы меньше $\frac{\delta}{\lambda}$, а у добавленного --- больше. То есть изменение увеличило вероятность в предположении альтернативной гипотезы, что мы сохраним нулевую гипотезу, что и требовалось доказать.

Иногда бывает, что один и тот же критерий позволяет сравнить нулевую гипотезу сразу со многими альтернативными. Например, пусть у нас есть серия подбрасываний монеты и p --- предполагаемая вероятность орла. Пусть мы хотим сравнить гипотезу, что p --- истинное значение с гипотезой, приписывающей значение $q > p$. По критерию Неймана-Пирсена мы должны рассмотреть для каждой выборки с k орлами разность $p^k(1-p)^{n-k} - \lambda q^k(1-q)^{n-k}$. Заметим, что нам надо просто выбрать при каких k мы сохраним нулевую гипотезу. Ясно, что с уменьшением k отношение $\left(\frac{p}{q}\right)^k \left(\frac{1-p}{1-q}\right)^{n-k}$ растёт, то есть нам надо взять некоторое количество наименьших k. То, сколько значений k надо взять, задаётся требованием, чтобы суммарная вероятность получить одно из этих значений как количество орлов при бросаниях монеты с вероятностью выпадения орла, равной p, была хотя бы $1 - \varepsilon$. Заметим, что это не зависит от q. Таким образом, для сравнения гипотезы о вероятности выпадения орла с гипотезой, предполагающей большую вероятность орла, надо потребовать, чтобы в выборке было не слишком много орлов (что ожидаемо), причём критическое количество не зависит от альтернативной гипотезы (если только $q > p$).

Ещё из статистических критериев широко известен критерий χ^2. Этот критерий применяется, когда у нас есть конечное количество (скажем, m) элементарных событий и проведено n испытаний. Пусть гипотетические вероятности элементарных событий $p_1, \ldots, p_m$, а при эксперименте они реализовались $\mu_1, \ldots, \mu_m$ раз, соответственно. Тогда оказывается, что $\sum\limits_k \frac{(\mu_k - np_k^0)^2}{np_i^0}$ распределено приблизительно как $\chi^2_{1-\epsilon}(m-1)$, где $\chi^2(r) = \eta_1^2 + \ldots + \eta_r^2$

--- сумма квадратов независимых нормальных (гауссовых) распределений. Критерий χ^2 отвергает гипотезу, если значение попадает в хвост распределения χ^2, суммарная вероятность которого меньше ε.

Тонкость с применением жёсткого критерия значимости проявляется в науке. Если 50 групп учёных ставит один и тот же эксперимент, то 2 из них получат несуществующий эффект со вполне допустимым уровнем значимости 5%. Так как статью про новый эффект опубликовать проще, все увидят статью про новое открытие и его воспроизведение --- и ничего более. Подчёркнуть проблему пытается своим существованием ``Журнал в поддержку нулевой гипотезы'', публикующий только статьи, в которых описаны качественно и аккуратно поставленные эксперименты, не продемонстрировавшие статистически значимых ранее неизвестных эффектов в наблюдавшихся явлениях.