§8. Равномерно наиболее мощные критерии

<u>*Лекция 13*</u>

Метод построения критериев заданного уровня α, который равномерно по всем альтернативным значениям параметра θ максимизирует мощность критерия, существенно опирается на следующее, почти очевидное утверждение, которое в теории проверки гипотез обычно называется *леммой Неймана–Пирсона.*

Рассмотрим вероятностную модель, состоящую всего из двух распределений P_0 и P_1, с общим носителем $\mathfrak{X}$ и функциями плотности $f_0(x)$ и $f_1(x)$, $x \in \mathfrak{X}$. По выборке $X^{(n)}$ проверяется простая гипотеза H_0 : выборка взята из распределения P_0 при простой альтернативе H_1 : выборке соответствует распределение P_1. Определим критическую функцию $\varphi^*(X^{(n)})$ как индикаторную функцию критической области

$$L(X^{(n)}) = \prod_{k=1}^{n} \frac{f_1(X_k)}{f_0(X_k)} > C.$$

Статистика L называется *статистикой отношения правдоподобия*, а критерий φ^* – *критерием отношения правдоподобия* или *критерием Неймана–Пирсона.* Критерий φ^* отвергает нулевую гипотезу, если правдоподобие альтернативы $f_{1,n}(X^{(n)}) = \prod_1^n f_1(X_k)$ в C раз превосходит правдоподобие нулевой гипотезы $f_{0,n}(X^{(n)}) = \prod_1^n f_0(X_k)$. Этот критерий обладает следующим замечательным свойством.

Теорема 8.1.*Критерий отношения правдоподобия φ^* является наиболее мощным критерием в классе всех критериев проверки простой гипотезы при простой альтернативе, размер которых не превосходит размера критерия φ^*. Если критерий φ^* имеет размер α, то он обладает наибольшей мощностью в классе всех критериев уровня α.*

Доказательство.. Пусть $\varphi = \varphi(X^{(n)})$ – любой другой критерий, размер которого

$$\mathbf{E}_0\, \varphi(X^{(n)}) \le \mathbf{E}_0\, \varphi^*(X^{(n)}). \tag{1}$$

Требуется показать, что тогда критерий φ^* имеет большую мощность, чем критерий φ, то есть $\mathbf{E}_1\, \varphi^*(X^{(n)}) \ge \mathbf{E}_1\, \varphi(X^{(n)})$.

Рассмотрим интеграл

$$\int\limits_{\mathfrak{X}^n} \left[\, \varphi^*(x^{(n)}) - \varphi(x^{(n)}) \right] \left[\, f_{1,n}(x^{(n)}) - C f_{0,n}(x^{(n)}) \right] d\mu_n(x^{(n)}) =$$

$$\mathbf{E}_1\,\varphi^*(X^{(n)}) - \mathbf{E}_1\,\varphi(X^{(n)}) - C\left[\,\mathbf{E}_0\,\varphi^*(X^{(n)}) - \mathbf{E}_0\,\varphi(X^{(n)})\,\right].$$

Достаточно показать, что этот интеграл неотрицателен, и тогда первое утверждение теоремы будет следовать из неравенства:

$$\mathbf{E}_1\,\varphi^*(X^{(n)}) - \mathbf{E}_1\,\varphi(X^{(n)}) - C\left[\,\mathbf{E}_0\,\varphi^*(X^{(n)}) - \mathbf{E}_0\,\varphi(X^{(n)})\,\right] \,\ge\, 0$$

которое влечет (см. (1))

$$\mathbf{E}_1\,\varphi^*(X^{(n)}) - \mathbf{E}_1\,\varphi(X^{(n)}) \,\ge\, C\left[\,\mathbf{E}_0\,\varphi^*(X^{(n)}) - \mathbf{E}_0\,\varphi(X^{(n)})\,\right] \,\ge\, 0.$$

Покажем, что функции $\varphi^*(x^{(n)}) - \varphi(x^{(n)})$ и $f_{1,n}(x^{(n)}) - Cf_{0,n}(x^{(n)})$, произведение которых интегрируется, одновременно положительны или отрицательны при любых $x^{(n)} \in \mathfrak{X}^n$. Действительно, если $\varphi^*(x^{(n)}) - \varphi(x^{(n)}) > 0$, то это влечет $\varphi^*(x^{(n)}) = 1$, поскольку критическая функция равна единице, если она не равна нулю. Но, по определению критерия отношения правдоподобия, равенство $\varphi^*(x^{(n)}) = 1$ возможно лишь в случае $f_{1,n}(x^{(n)}) - Cf_{0,n}(x^{(n)}) > 0$. Точно также устанавливается, что неравенство $\varphi^*(x^{(n)}) - \varphi(x^{(n)}) < 0$ влечет $f_{1,n}(x^{(n)}) - Cf_{0,n}(x^{(n)}) < 0$.

Итак, критерий φ^* наиболее мощен в классе всех критериев, размер которых не превосходит размера φ^*. Если же $\mathbf{E}_0\,\varphi^*(X^{(n)}) = \alpha$, то это утверждение, очевидно, влечет его наибольшую мощность в классе всех критериев уровня α.

Применение этой теоремы к построению равномерно наиболее мощных критериев мы проиллюстрируем на одном частном примере, из которого будет виден общий подход к данной задаче.

Пример 8.1. *Проверка надежности при показательном распределении долговечности.* В примере 3.3 мы рассматривали проблему оценки надежности изделия с показательным распределением долговечности. Напомним, случайная величина X, реализация x которой соответствует промежутку времени от начала работы до момента отказа некоторого изделия, называется долговечностью, и по функции распределения $F(x)$, $x \ge 0$ случайной величины X можно рассчитать надежность $H(t)$ изделий, соответствующую гарантийному времени $t:\ H(t) = P(X \ge t) = 1 - F(t)$.

Пусть долговечность X распределена по показательному закону с функцией распределения $F(x\,|\,\theta) = 1 - \exp\{-x/\theta\}$, значение параметра θ которой не известно. Мы должны удостовериться, что надежность выпускаемых изделий достаточно высока: $H(t) \geq P_0$, где P_0 – наименьшая допустимая доля изделий, которые должны прослужить гарантийный срок t.

Это типичная задача проверки гипотез, решение которой начинается с определения нулевой гипотезы H_0. При этом следует помнить, что в статистическом критерии контролируется вероятность отклонения H_0, когда она в действительности верна. В нашей конкретной проблеме спецификация нулевой гипотезы во многом зависит от того, что повлечет за собой отказ изделия. Если мы выпускаем бытовые приборы, то отказ изделия до гарантийного скрока t повлечет издержки на ремонт, которые могут быть незначительными по сравнению со стоимостью изделия. В таком случае естественно выбрать в качестве нулевой гипотезы утверждение о надежности изделий – отклонив эту гипотезу, когда она верна, мы потеряем дорогостоящую продукцию, ремонт которой нам обошелся бы значительно дешевле, чем ее уничтожение или продажа по бросовой цене. Если же отказ изделия приводит к катастрофическим последствиям, например, к гибели людей, то здесь рассуждать нечего, и за нулевую гипотезу следует брать утверждение о “ненадежности”. Отклонив такую гипотезу, когда она в действительности верна, мы столкнемся с неприемлемо большой долей отказов до истечения гарантийного срока, и поэтому риск от принятия “плохих” изделий должен быть контролируем. Остановимся на этом варианте и приступим к построению равномерно наиболее мощного критерия проверки гипотезы “ненадежности” $H_0: \ H(t) < P_0$ при альтернативе $H_1: \ H(t) \geq P_0$, когда $H(t) = \exp\{-t/\theta\}$.

В терминах значений параметра θ нулевая гипотеза принимает вид $H_0: \ \theta < \theta_0 = -t/\ln\alpha$. Зафиксируем некоторое альтернативное значение $\theta_1 > \theta_0$, и рассмотрим задачу проверки простой гипотезы $H_0': \ \theta = \theta_0$ при простой альтернативе $H_1': \ \theta = \theta_1$. Наиболее мощный критерий проверки простой гипотезы при простой альтернативе имеет критическую область вида (см. теорему 8.1)

$$L(X^{(n)}) = \prod_{k=1}^{n} \frac{f_1(X_k)}{f_0(X_k)} = \frac{\theta_0}{\theta_1} \exp\left\{\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) \textstyle\sum_1^n X_k\right\} > C,$$

где критическая константа C определяется по заданному уровню значимости α из условия $P_{\theta_0}\left(L(X^{(n)}) > C\right) \leq \alpha$. Поскольку статистика $T_n = \sum_1^n X_k$ имеет гамма-распределение $\mathrm{G}(n, \theta_0)$, то для определения

C в последнем неравенстве следует положить знак равенства. Кроме этого, статистика отношения правдоподобия $L(X^{(n)})$ есть монотонная функция статистики T_n, поэтому критическую область $L(X^{(n)}) > C$ можно записать в эквивалентной форме $T_n > C$ и находить новое C из равенства $P_{\theta_0}(T_n > C) = 1 - G_n(C/\theta_0) = \alpha$ (собственно говоря, нам все равно, какое C определять, но на практике, вне сомнения, удобнее иметь дело с критической областью $T_n > C$).

Итак, $C(\alpha) = \theta_0 \cdot G_n^{-1}(1-\alpha)$, где $G_n^{-1}(\cdot)$ – квантиль стандартного гамма-распределения $\mathrm{G}(n,1)$, и критерий $\varphi^*(X^{(n)}) = \mathrm{I}_{\{T_n > C(\alpha)\}}(X^{(n)})$ заданного размера α является наиболее мощным в классе всех критериев уровня α, проверяющих гипотезу H_0' при альтернативе H_1'. Это означает, что для любого другого критерия φ с $\mathbf{E}_{\theta_0}\,\varphi(X^{(n)}) \le \alpha$ выполняется неравенство

$$\mathbf{E}_{\theta_1}\,\varphi(X^{(n)}) \le \mathbf{E}_{\theta_1}\,\varphi^*(X^{(n)}). \tag{2}$$

Но критерий φ^* не зависит от выбора альтернативного значения θ_1 параметра θ – критическая константа $C(\alpha) = \theta_0 \cdot G_n^{-1}(1-\alpha)$! Следовательно, неравенство (2) справедливо при любых $\theta_1 > \theta_0$, и мы приходим к заключению, что *критерий φ^* есть равномерно наиболее мощный критерий в классе всех критериев уровня α, проверяющих простую гипотезу $H_0' : \ \theta = \theta_0$ при сложной альтернативе $H_1 : \ \theta > \theta_0$.*

Далее, функция мощности критерия φ^*, как критерия различения исходных сложных гипотез $H_0 : \ \theta < \theta_0$ и $H_1 : \ \theta \ge \theta_0$, равна $m(\theta) = \mathbf{E}_\theta\,\varphi^*(X^{(n)}) = P_\theta(T_n > C(\alpha) = 1 - G_n(G^{-1}(1-\alpha)\theta_0/\theta)$, $\theta > 0$. Это – возрастающая функция θ, поэтому максимум вероятности ошибки первого рода (размер критерия) равен $m(\theta_0) = 1 - G_n\left(G_n^{-1}(1-\alpha)\right) = \alpha$. Таким образом, *критерий φ^* есть критерий размера α проверки гипотезы H_0 при альтернативе H_1, обладающий равномерно наибольшей мощностью в классе всех критериев φ с ограничением* $\mathbf{E}_{\theta_0}\,\varphi(X^{(n)}) = \alpha$. Но в таком случае он будет равномерно наиболее мощным и в более узком классе критериев уровня α, то есть критериев φ, удовлетворяющих ограничению $\mathbf{E}_\theta\,\varphi(X^{(n)}) \le \alpha$ при любом $\theta < \theta_0$.

Более того, нетрудно убедиться, что критерий φ^* обладает минимальной вероятностью ошибки первого рода $\alpha(\theta) = m(\theta), \ \theta \le \theta_0$ в классе всех критериев уровня α. Для этого достаточно поменять местами нулевую гипотезу и альтернативу и выбрать уровень значимости, равный $1-\alpha$.

В этом примере построение равномерно наиболее мощного критерия стало возможным благодаря особому свойству статистической структуры показательного распределения: *статистика* $L(X^{(n)})$ *отношения правдоподобия есть монотонная функция статистики* $T_n = \sum_1^n X_k$. Это – частный случай статистических структур, обладающих достаточной статистикой T, ибо в силу теоремы факторизации у таких структур $L(X^{(n)}) = g_{\theta_1}(T)/g_{\theta_0}(T)$ зависит от $X^{(n)}$ только через значения $T(X^{(n)})$. Дополнительное свойство монотонности отношения правдоподобия относительно T обеспечивает существование и возможность конструктивного построения равномерно наиболее мощного критерия, причем критическая область такого критерия обязательно имеет вид $T > C$ или $T < C$. Например, критерий $\sum_1^n X_k > C$ при соответствующем выборе C по заданному уровню значимости α будет равномерно наиболее мощным критерием в классе всех критериев уровня α проверки гипотезы $\theta < \theta_0$ при альтернативе $\theta \geq \theta_0$, когда θ есть среднее значение нормального распределения (дисперсия предполагается известной) или параметр масштаба гамма-распределения (параметр формы известен). Но если θ – параметр таких распределений, как двухточечное или Пуассона, то критерий φ^* с критической областью $\sum_1^n X_k > C$ обладает равномерно наибольшей мощностью только в классе тех критериев, размер которых не больше размера φ^*.

Другие критерии, которые мы рассматривали в предыдущем параграфе, также обладают свойством равномерной наибольшей мощности, и при доказательстве этого также используется лемма Неймана–Пирсона, но методика доказательства совершенно другая и требует разработки методов построения критериев, обладающих свойством инвариантности – независимости от мешающих параметров. Но это уже совсем другая область теории проверки гипотез, поговорить о которой у нас не хватает времени. Я лучше расскажу вам о некоторых дополнительных ухищрениях в практических применениях статистических критериев, которые позволяют с большей степенью наглядности оценить степень согласия проверяемой гипотезы с выборочными данными.

Все рассматриваемые нами критерии заданного уровня α обладают тем свойством, что их критические области можно записать в виде $T(X^{(n)}) > C(\alpha)$, где T – некоторая статистика, характеризующая расхождение выборочных данных с предполагаемыми значениями параметра. Увеличение уровня значимости α приводит к уменьшению $C(\alpha)$, и мы получаем систему вложенных друг в друга критических областей. Это замечательное свойство наших критериев позволяет несколько изменить методологию их практического использования. До сих пор мы фиксировали уровень значимости α, находили по нему критическую константу $C(\alpha)$ и сравнивали ее с выборочным значением $t = T(x^{(n)})$ статистики $T = T(X^{(n)})$. Поступим теперь следующим образом. Получив выборочные данные $x^{(n)}$, вычислим значение $t = T(x^{(n)})$ и рассмотрим критерий $T(X^{(n)}) > t$. Размер такого критерия $\alpha_{кр.} = P_0\left(T(X^{(n)}) > t\right)$ называется *критическим уровнем значимости*, который трактуется как вероятность получить столь же большие расхождения между выборочными данными и нулевой гипотезой, как и для выборочных данных $x^{(n)}$.

Естественно, мы по-прежнему можем работать с заданным уровнем значимости α, отклоняя нулевую гипотезу, если $\alpha_{кр.} < \alpha$, и принимая ее в противном случае. Кстати, принимая гипотезу, не следует утверждать, что она верна. На этот счет существует более деликатное выражение: "выборочные данные согласуются с выдвинутой гипотезой," ибо, как говорил один из создателей математической статистики сэр Д.Фишер, "гипотезы не проверяютя, а разве лишь отвергаются". Так вот, в свете этого высказывания более разумно просто сообщать полученный критический уровень значимости, сопровождая его следующим комментарием, который можно считать международным статистическим стандартом. Если $\alpha_{кр.} \leq 0.01$, то говорят, что расхождение между гипотезой и выборочными данными *высоко значимо*, если $0.01 < \alpha_{кр.} \leq 0.05$, то просто – *значимо*, если же $0.05 < \alpha_{кр.} \leq 0.10$ – *почти значимо*, и в случае $\alpha_{кр.} > 0.10$ – *не значимо*. Заметим также, что в некоторых применениях критериев значимости (особенно, в медицине) $\alpha_{кр.}$ называют *достоверностью*. Существуют и другие, совершенно фантастические названия $\alpha_{кр.}$, которые я не буду здесь приводить в силу их крайне неприличного звучания.

Поговорим теперь об оптимальных свойствах доверительных границ, соответствующих равномерно наиболее мощным критериям. Рассмотрим только случай верхней доверительной границы $\overline{\theta}_n = \overline{\theta}_n(X^{(n)})$.

Определение 8.1 Верхняя $(1-\alpha)$-доверительная граница $\overline{\theta}_n$ называется *равномерно наиболее точной*, если она равномерно по всем θ и θ', удовлетворяющим неравенству $\theta' > \theta$, минимизирует вероятность $P_\theta\left(\overline{\theta}_n(X^{(n)}) \geq \theta'\right)$.

Таким образом, в случае равномерно наиболее точной границы $\overline{\theta}_n$ интервал $(-\infty; \overline{\theta}_n]$ с заданной вероятностью $1-\alpha$ накрывает истинное значение параметра θ, но он с минимальной вероятностью накрывает любые значения θ, лежащие правее истинного.

Если мы проверяем гипотезу $H: \ \theta = \theta_0$ при альтернативе $K(\theta_0): \ \theta < \theta_0$, и область принятия $A(\theta_0)$ равномерно наиболее мощного критерия размера α обладает тем свойством, что подмножество $\Delta_n(x^{(n)}) = \{\theta: \ x^{(n)} \in A(\theta)\}$ параметрического пространства $\Theta \subseteq \mathrm{R}$ есть интервал $(-\infty : \overline{\theta}_n(x^{(n)})]$, то $\overline{\theta}_n(X^{(n)})$ есть равномерно наиболее точная верхняя $(1-\alpha)$-доверительная граница. Все объясняется довольно просто: вероятность $P_\theta\left(\overline{\theta}_n(X^{(n)}) \geq \theta'\right) = P_\theta\left(X^{(n)} \in A(\theta')\right)$ есть вероятность ошибки второго рода у критерия проверки гипотезы $H: \ \theta = \theta'$ при альтернативе $K(\theta'): \ \theta < \theta'$. Равномерно наиболее мощный критерий естественно обладает равномерно минимальной вероятностью ошибки второго рода. Все построенные нами в §6 доверительные границы обладают оптимальными свойствами с точки зрения малой вероятности накрытия тех значений параметра, которые не соответствуют истине.