§7. Статистическая проверка гипотез (критерии значимости)

Лекция 11

В приложениях математической статистики существует обширный класс задач, в которых требуется проверить истинность некоторого высказывания относительно исследуемого объекта или выбрать одно из альтернативных решений, которое определит дальнейшее поведение статистика по отношению к этому объекту. Например, при аттестации партии дизельного топлива по общему содержанию серы мы должны не только дать точечную оценку данной характеристики топлива, но и принять решение о качестве выпускаемого продукта, которое повлечет за собой одно из следующих действий – или отослать топливо потребителю, или произвести дополнительную очистку топлива от вредных примесей. Точно так же в примере 1.2 мы строили статистическое правило, позволяющее принять одно из двух решений относительно нового лечебного препарата – или признать его эффективным и внедрить в лечебную практику, или запретить его дальнейшее использование. В исследованиях, подобных опытам Менделя, часто надо проверить гипотезу относительно предполагаемого значения вероятности наследования доминантного признака. Селекционер, работающий над получением нового вида пшеницы, должен подкрепить свое заключение о превосходстве нового вида над тем, который уже используется в сельскохозяйственной практике, с помощью сопоставления данных об урожайности этих видов. И так далее, и тому подобное, – вы сами можете привести примеры таких задач по выбору одного из ряда альтернативных решений.

В нашем курсе математической статистики мы рассмотрим задачи, связанные только с выбором одного из двух решений. Пусть мы высказываем некоторое суждение (или предпринимаем действие) об исследуемом объекте, и пусть d_0 – решение об истинности этого суждения, в то время как d_1 – решение о его ложности. Таким образом, пространство решений $\mathcal{D}$ в данной статистической проблеме состоит из точек: $\mathcal{D} = \{d_0, d_1\}$.

Для выбора одного из решений мы наблюдаем случайную выборку $X^{(n)}$ из некоторого распределения P_θ, значение параметра θ которого нам неизвестно. Пусть Θ – область возможных значений θ, которую мы назвали параметрическим пространством. В соответствии с принятой нами в §1 идеологией статистического вывода мы сопоставляем каждому решению $d \in \mathcal{D}$ определенное подмножество Θ_d пространства Θ, то есть интерпретируем каждое решение в терминах высказываний об истинном значении параметра θ. В нашей статистической проблеме выбора одного из двух решений положим $\Theta_i = \Theta_{d_i},\ i = 0, 1$, и введем ряд понятий и определений, используемых при решении этой проблемы.

Утверждение $H_0 : \ \theta \in \Theta_0$ называется *нулевой гипотезой*, а утверждение $H_1 : \theta \in \Theta_1$ – *альтернативной гипотезой* или (коротко) *альтернативой*. Гипотеза H_i называется *простой*, если соответствующее Θ_i состоит из одной точки параметрического пространства Θ; в противном случае H_i называется *сложной* гипотезой; $i = 0, 1$. Так, в примере 1.2 с испытанием нового лечебного препарата параметр θ означал вероятность успешного лечения каждого пациента, и нулевая гипотеза $H_0 : \theta = 1/2$ о "нейтральности" препарата есть простая гипотеза, в то время как альтернативная гипотеза $H_1 : \theta > 1/2$ об его эффективности – сложная гипотеза.

Правило, по которому принимается или отвергается нулевая гипотеза H_0, называется *критерием*. Иногда добавляется – *критерий согласия* (с нулевой гипотезой), особенно, когда альтернатива H_1 определена не совсем четко и под H_1 подразумевается "все остальное". В случае полного равноправия гипотез говорят о критерии *различения гипотез*. Критерий определяется заданием особого подмножества S выборочного пространства $\mathcal{X}^n$, которое называется *критической областью*: если выборочные данные $x^{(n)}$ попадают в эту область, то нулевая гипотеза H_0 отклоняется и принимается альтернативное решение – справедлива H_1. Область $A = S^c = \mathcal{X}^n \setminus S$ называется *областью принятия* нулевой гипотезы. Нам будет удобно проводить спецификацию критической области в виде ее индикаторной функции $\varphi = \varphi(X^{(n)})$, которая называется *критической функцией* или, поскольку она определяет статистическое правило проверки гипотезы, просто *критерием*. Итак, функция $\varphi(X^{(n)})$ есть бинарная случайная величина, принимающая значение 1, если произошло событие $X^{(n)} \in S$, и значение 0, если произошло противоположное событие $X^{(n)} \in A$. Понятно, что математическое ожидание $\mathbf{E}\varphi(X^{(n)})$ означает вероятность отклонения гипотезы H_0.

В рассматриваемой статистической проблеме величина риска, связанная с отклонением верной гипотезы, обычно соотносится с функцией потерь типа $1-0$: потери считаются равными 1, если принята гипотеза H_i, а в действительности $\theta \in \Theta_{1-i}$, $i = 0, 1$; если же принята H_i и $\theta \in \Theta_i$, $i = 0, 1$, то потери полагаются равными нулю. Легко видеть, что величина риска при любом значении параметра θ может быть определена с помощью функции $m(\theta) = \mathbf{E}_\theta \varphi(X^{(n)}) = P_\theta(X^{(n)} \in S)$, которая называется *функцией мощности* критерия φ. Эта функция указывает, как часто мы отклоняем нулевую гипотезу, когда θ – истинное значение параметра, и хорошим следует считать тот критерий, у которого функция $m(\theta)$ принимает близкие к нулю значения в области Θ_0 и близкие к единице – в области Θ_1. В связи с этим вводятся две компоненты функции риска: $\alpha(\theta) = m(\theta)$ при $\theta \in \Theta_0$ и $\beta(\theta) = 1 - m(\theta)$ при $\theta \in \Theta_1$. Функция $\alpha(\theta)$, $\theta \in \Theta_0$ называется *вероятностью ошибки первого рода* – она указывает относительную частоту отклонения гипотезы H_0, когда она в действительности верна $(\theta \in \Theta_0)$. Функция $\beta(\theta)$, $\theta \in \Theta_1$ называется *вероятностью ошибки второго рода* – она указывает относительную частоту принятия гипотезы H_0, когда она ложна (верна альтернативная гипотеза $H_1 : \ \theta \in \Theta_1$). Заметим, что функция мощности $m(\theta)$ в области Θ_1 трактуется как вероятность отклонения гипотезы H_0, когда в действительности выбор идет из распределения с альтернативным значением $\theta \in \Theta_1$, и поэтому часть $m(\theta)$ при $\theta \in \Theta$ называется *мощностью* критерия φ.

Легко понять, что при фиксированном объеме наблюдений n невозможно одновременно минимизировать вероятности обеих ошибок, – для уменьшения вероятности ошибки первого рода $\alpha(\theta) = P_\theta(X^{(n)} \in S)$, $\theta \in \Theta_0$, необходимо уменьшить критическую область S, что приведет к увеличению области A принятия нулевой гипотезы и, следовательно, к увеличению вероятности ошибки второго рода $\beta(u) = P_u(X^{(n)} \in A)$, $u \in \Theta_1$. Здесь возникает такая же ситуация, что и в проблеме построения оценки параметра θ с равномерно минимальным риском, – такие оценки существуют только в определенном классе статистических правил, например, в классе несмещенных оценок. Однако, даже и помимо задачи проверки гипотез с минимальной вероятностью ошибки, и намного раньше создания общей теории наиболее мощных критериев в статистической практике сложился следующий подход к управлению риском критерия.

Предположим, что отклонение гипотезы H_0, когда она в действительности верна, приводит к более тяжким последствиям, чем ее принятие при справедливости альтернативы. В таком случае мы заинтересованы в первую очередь контролировать вероятность ошибки первого рода. С этой целью заранее фиксируется (выбирается) некоторый уровень α, выше которого вероятность ошибки первого рода не допустима, и критическая область S (критерий φ) определяется таким образом, что $\alpha(\theta) \leq \alpha$, каково бы ни было $\theta \in \Theta_0$. Это ограничение α на вероятность ошибки первого рода называется *уровнем значимости*, а сам критерий φ, для которого выполняется это ограничение, – *критерием уровня* α. Наибольшее значение вероятности ошибки первого рода

$$\overline{\alpha} = \sup_{\theta \in \Theta_0} \alpha(\theta)$$

называется *размером* критерия φ, и если $\overline{\alpha} = \alpha$, то говорят о *критерии* φ *размера* α.

В этом выборе ограничения именно на вероятность ошибки первого, а не второго рода проявляется типичная асимметрия в практической ценности гипотезы и альтернативы. Например, если проверяется эффективность нового лекарственного препарата, то нулевой гипотезе должно соответствовать решение о его неэффективности, ибо, отклонив эту гипотезу, когда она верна, мы внедрим в лечебную практику бесполезное или вредное лекарство, что приведет к более тяжким последствиям, чем отклонение в действительности эффективного препарата. Но если мы ищем золото, анализируя состав кернов при бурении предполагаемого месторождения, то естественно принять за нулевую гипотезу утверждение о наличии золота, ибо отклонив ее, когда она верна, мы потеряем намного больше, чем стоимость нескольких дополнительных анализов, удостоверяющих, что золото в разбуренной местности отсутствует.

Следует также обратить особое внимание на общую методологию проверки гипотез, отражаемую в выборе малого значения уровня α. Если наши выборочные данные попадают в область S с исключительно малой вероятностью, то естественно предположить, что то утверждение, которое привело к этому маловероятному событию, не соответствует истине и отклонить его. Поступая таким образом, мы будем терять в действительности верную гипотезу H_0 крайне редко

– не более, чем в $100\alpha\%$ случаев.

Простейший метод построения критериев значимости состоит в использовании состоятельных оценок тестируемого параметра θ. Рассмотрим простейший случай: θ – скалярный параметр, вероятностная модель не содержит других (мешающих) параметров и проверяется простая гипотеза $H_0: \ \theta = \theta_0$ при альтернативе $H_1: \ \theta \neq \theta_0$, где θ_0 – некоторое, априори фиксированное значение параметра θ (например, в опытах Менделя проверяется гипотеза: вероятность θ наследования доминантного признака равна $\theta_0 = 3/4$). Если $\hat{\theta}_n = \hat{\theta}_n(X^{(n)})$ – состоятельная оценка θ, дисперсия которой стремится к нулю при $n \to \infty$ как $O(1/\sqrt{n})$, то естественно определить критическую область посредством неравенства $\sqrt{n}|\,\hat{\theta}_n(X^{(n)}) - \theta_0\,| > C$. Вероятность ошибки первого рода такого критерия $\alpha(\theta_0, C) = P_{\theta_0}(\sqrt{n}|\,\hat{\theta}_n(X^{(n)}) - \theta_0\,| > C)$, приравнивая которую заданному уровню значимости α находим *критическую константу* $C = C(\alpha)$ как квантиль распределения случайной величины $\sqrt{n}|\,\hat{\theta}_n(X^{(n)}) - \theta_0\,|$; такой выбор C приводит к критерию уровня α. Если $\theta\,(\neq \theta_0)$ – некоторое альтернативное значение параметра, то, в силу состоятельности оценки, $\sqrt{n}|\,\hat{\theta}_n(X^{(n)}) - \theta_0\,| \underset{P}{\to} \infty$, и поэтому вероятность ошибки второго рода $\beta(\theta) = P_\theta(\sqrt{n}|\,\hat{\theta}_n(X^{(n)}) - \theta_0\,| > C(\alpha) \to 0$, когда $n \to \infty$. Таким образом, мы получаем критерий заданного уровня α, обладающий к тому же свойством *состоятельности* – его вероятность ошибки второго рода стремится к нулю при неограниченном возрастании объема выборки n.

Сформулируем теперь основную задачу теории статистической проверки гипотез: *требуется найти такой критерий φ уровня α, который равномерно по всем $\theta \in \Theta_1$ максимизирует мощность $m(\theta)$ или, что то же, равномерно по $\theta \in \Theta_1$ минимизирует вероятность ошибки второго рода $\beta(\theta)$.* Мы укажем метод построения таких *равномерно наиболее мощных критериев* заданного уровня α в следующем параграфе, а пока обратимся к иллюстрациям введенных понятий и построению наиболее часто используемых на практике критериев, касающихся проверки гипотез о значениях параметров нормального распределения.

1^0. Проверка гипотезы о величине среднего значения нормального распределения при известной дисперсии. Рассмотрим сначала наиболее часто встречающуюся в практических применениях математической статистики задачу проверки сложной гипотезы $H_0: \ \mu \le \mu_0$ при сложной альтернативе $H_1: \ \mu > \mu_0$ о среднем значении μ нормального (μ, σ^2) распределения при известном значении дисприсии σ^2. Выборочное среднее $\overline{X} = n^{-1}\sum_1^n X_k$ есть оптимальная оценка неизвестного значения μ, и поэтому, в соответствии с только что предложенным методом построения состоятельных критериев, рассмотрим критерий, отвергающий нулевую гипотезу H_0, когда $\sqrt{n}(\overline{X} - \mu_0) > C$, или, что то же, $\overline{X} > C$, поскольку значения μ_0 и n фиксированы и известны. Постоянная C должна выбираться по заданному уровню значимости α, ограничивающему максимальное значение вероятности ошибки первого рода.

Так как при выборе из нормального (μ, σ^2) распределения статистика $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, то функция мощности этого критерия

$$m(\mu) = P_\mu(\overline{X} > C) = 1 - \Phi\left(\frac{C-\mu}{\sigma}\sqrt{n}\right) = \Phi\left(\frac{\mu - C}{\sigma}\sqrt{n}\right).$$

Легко видеть, что $m(\mu)$ – строго возрастает с ростом μ, так что размер критерия

$$\overline{\alpha} = \max_{\mu \le \mu_0} m(\mu) = m(\mu_0) = 1 - \Phi\left(\frac{C-\mu_0}{\sigma}\sqrt{n}\right).$$

Приравнивая размер критерия уровню значимости α, находим критическое значение $C(\alpha) = \mu_0 + \Phi^{-1}(1-\alpha)\sigma/\sqrt{n}$.

Вероятность ошибки второго рода нашего критерия размера α

$$\beta(\mu) = P_\mu(\overline{X} \le C(\alpha)) = \Phi\left(\frac{C(\alpha)-\mu}{\sigma}\sqrt{n}\right) =$$

$$\Phi\left(\frac{\mu_0 - \mu}{\sigma}\sqrt{n}\Phi^{-1}(1-\alpha)\right), \ \mu > \mu_0, \tag{1}$$

убывает с ростом μ по мере ее отхода от граничного значения μ_0. Наибольшее значение $\beta(\mu)$ достигается в точке $\mu = \mu_0$ и равно $1 - \alpha$. Это значение не зависит от размера выборки n, и поэтому требуются дополнительные соображения при планировании объема наблюдений. Обычно используется метод введения так называемой *зоны безразличия* – интервала (μ_0, μ_1), который выбирается из тех соображений, что при истинном значении $\mu \in (\mu_0, \mu_1)$ принятие нулевой гипотезы H_0 не приводит к слишком тяжелым последствиям. Однако при истинном $\mu \ge \mu_1$ вероятность принятия H_0 должна быть под контролем и не превосходить некоторого предписанного значения β. Это обстоятельство позволяет спланировать объем выборки n, определив его из неравенства $\beta(\mu_1) \le \beta$. Используя формулу (1) для $\beta(\mu)$, находим, что *объем выборки* $n = n(\alpha, \beta, \mu_0, \mu_1)$, *необходимый для различения гиптез* $\mu \le \mu_0$ *и* $\mu \ge \mu_1$ *с заданными ограничениями* α *и* β *на вероятности ошибок первого и второго рода,* равен наименьшему целому n, удовлетворяющему неравенству

$$n \ge \frac{\left[\,\Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta)\,\right]^2}{(\mu_1 - \mu_0)^2}\,\sigma^2.$$

Аналогичным методом строится критерий для проверки простой гипотезы $H_0: \ \mu = \mu_0$ при сложной альтернативе $H_1: \ \mu \ne \mu_0$. В этой задаче естественно определить критическую область посредством неравенства $|\,\overline{X} - \mu_0\,| > C$. Функция мощности такого критерия

$$m(\mu) = 1 - \left[\,\Phi\left(\frac{C + \mu_0 - \mu}{\sigma}\sqrt{n}\right) - \Phi\left(\frac{-C + \mu_0 - \mu}{\sigma}\sqrt{n}\right)\right]$$

строго убывает при $\mu < \mu_0$, возрастает при $\mu > \mu$ и при $\mu = \mu_0$ равна вероятности ошибки первого рода. Таким образом, критическая константа $C = C(\alpha)$ определяется по заданному уровню значимости α из уравнения

$$m(\mu_0) = 1 - \left[\,\Phi\left(\frac{C}{\sigma}\sqrt{n} - \right)\Phi\left(\frac{-C}{\sigma}\sqrt{n}\right)\right] = 2\left[1 - \Phi(\frac{C}{\sigma}\sqrt{n}\right] = 1 - \alpha,$$

откуда $C(\alpha) = \Phi^{-1}(1 - \alpha/2)\sigma/\sqrt{n}$.

Лекция 12

2^0. Проверка гипотезы о величине дисперсии нормального распределения при неизвестном среднем значении. Это типичная задача контроля за величиной случайной ошибки в параллельных наблюдениях некоторой характеристики исследуемого объекта. Так как превышение характеристики случайной погрешности σ над некоторым номиналом σ_0 в случае, когда мы утверждаем $\sigma \le \sigma_0$, влечет более серьезные последствия, чем неоправданные претензии к слишком большому разбросу в данных, то следует принять за нулевую гипотезу $\sigma > \sigma_0$. Проверка этой гипотезы проводится при естественной альтернативе $H_1 : \ \sigma \le \sigma_0$, причем мы не знаем значения мешающего параметра μ – среднего значения нормального распределения, из которого производится выбор.

Как нам известно, выборочная дисперсия $S^2 = n^{-1}\sum_1^n (X_k - \overline{X})^2$ есть состоятельная оценка σ^2, ее распределение не зависит от μ, а случайная величина nS^2/σ^2 имеет хи-квадрат распределение с $n-1$ степенью свободы. Таким образом, разумно рассмотреть критерий с критической областью $nS^2 < C$. Функция мощности такого критерия

$$m(\sigma) = P_{\mu,\sigma}\left(\frac{nS^2}{\sigma^2} \le \frac{C}{\sigma^2}\right) = K_{n-1}\left(\frac{C}{\sigma^2}\right)$$

монотонно убывает с ростом σ, поэтому наибольшее значение вероятности ошибки первого рода достигается при $\sigma = \sigma_0$, и критическое значение $C(\alpha)$ критерия требуемого размера α определяется из уравнения $K_{n-1}\left(C\sigma_0^{-2}\right) = \alpha$. Итак, $C(\alpha) = \sigma_0^2 K_{n-1}^{-1}(\alpha)$; вероятность ошибки второго рода

$$\beta(\sigma) = P_{\mu,\sigma}\left(nS^2 > C(\alpha)\right) = 1 - K_{n-1}\left(\frac{\sigma_0^2}{\sigma^2}K_{n-1}^{-1}(\alpha)\right), \ \sigma \le> \sigma_0,$$

монотонно убывает по мере отхода истинного значения σ от номинала σ_0.

3^0. Проверка гипотезы о величине среднего значения нормального распределения при неизвестной дисперсии (одновыборочный критерий Стьюдента). Вы, наверное, обратили внимание, что при построении критериев значимости мы по существу используем методы построения доверительных множеств? Это, действительно, так – между задачами доверительной оценки и проверки гипотез существует много общего, и, решив одну задачу, мы сразу же получаем решение другой. В конце этого параграфа мы формализуем этот параллелизм, а пока будем использовать его на интуитивном уровне: предлагается использовать для проверки гипотез о среднем значении нормального распределения статистику Стьюдента.

Рассмотрим сначала задачу проверки сложной гипотезы $H_0 : \ \mu \le \mu_0$ при сложной альтернативе $H_1 : \ \mu > \mu_0$. Так как выборочное среднее $\overline{X}$ есть состоятельная оценка значения μ, то статистика Стьюдента

$$T = \frac{\overline{X} - \mu_0}{S}\sqrt{n-1}$$

опосредственно, через выборочные данные, характеризует удаленность истинного среднего значения μ от гранцы μ_0, разделяющей гипотезу и альтернативу. Поэтому предлагается отвергать нулевую гипотезу $\mu \leq \mu_0$, если $T > C$, выбирая C, как обычно, по заданному уровню значимости α. Для решения последней задачи необходимо исследовать поведение функции мощности $m(\mu; \sigma) = P_{\mu,\sigma}(T > C)$ критерия $T > C$. Если мы покажем, что $m(\mu : \sigma)$ есть монотонно возрастающая функция аргумента μ при любом фиксированном значении аргумента σ, то наибольшее значение вероятности ошибки первого рода $\alpha(\mu; \sigma) = m(\mu : \sigma)$, $\mu \leq \mu_0$ при каждом фиксированном σ будет достигаться в точке $\mu = \mu_0$. Следовательно, размер критерия в таком случае будет равен (см. пункт 4^0 предыдущего параграфа) $\overline{\alpha} = m(\mu_0; \sigma) = P_{\mu_0,\sigma}(T > C) = 1 - S_{n-1}(C)$, где $S_\nu(\cdot)$ – функция распределения Стьюдента с ν степенями свободы. Таким образом, мы получим свободный от неизвестного значения σ критерий $T > C(\alpha)$ требуемого размера α с критической константой $C(\alpha) = S_{n-1}^{-1}(1 - \alpha)$. Это и есть то статистическое правило, которое обычно называется *критерием Стьюдента* или *t-критерием.*

Покажем теперь, что вероятность (функция мощности) $P_{\mu,\sigma}(T > C)$ монотонно возрастает с ростом μ при любых фиксированных значениях σ и C. С этой целью представим статистику T в следующем виде:

$$T = \frac{\overline{X} - \mu}{S}\sqrt{n-1} + \frac{\mu - \mu_0}{\sigma} \cdot \frac{\sigma}{S}\sqrt{n-1}.$$

Если μ – среднее значение нормального распределения, из которого происходит выбор, то первое слагаемое в этом представлении есть стьюдентовская случайная величина с $n-1$ степенью свободы. Второе слагаемое есть произведение параметрической функции $\Delta(\mu) = (\mu - \mu_0)\sqrt{n-1}/\sigma$ на положительную случайную величину σ/S, распределение которой не зависит от μ и σ. При фиксированном σ функция $\Delta(\mu)$ возрастает с ростом μ и при этом все второе слагаемое возрастает, что влечет увеличение вероятности события перескока статистикой T порога C, то есть вероятности события $T > C$.

Итак, мы построили критерий проверки *односторонней* гипотезы $\mu \le \mu_0$ при односторонней альтернативе $\mu > \mu_0$. Функция мощности этого критерия зависит от μ и σ только через параметрическую функцию $\Delta = (\mu - \mu_0)\sqrt{n-1}/\sigma$, которая называется *параметром нецентральности*. Распределение статистики $T = (\overline{X} - \mu_0)\sqrt{n-1}/S$ при произвольных μ и σ, через которое выражается функция мощности критерия Стьюдента, называется *нецентральным* распределением Стьюдента с $n-1$ степенью свободы; таблицы этого распределения, зависящего от параметра нецентральности Δ, можно найти в ТМС.

Понятно, что построение критерия проверки простой гипотезы $\mu = \mu_0$ при *двусторонней* (сложной) альтернативе $\mu \ne \mu_0$ не вызывает принципиальных затруднений. Это критерий с критической областью $|\,T\,| > C$, где критическая константа $C = C(\alpha) = S_{n-1}^{-1}(1 - \alpha/2)$.

4^0. Сравнение средних значений двух нормальных распределений с общей неизвестной дисперсией (двухвыборочный критерий Стьюдента). Пусть X и Y – независимые случайные величины, причем $X \sim \mathcal{N}(\mu_1, \sigma^2)$, а $Y \sim \mathcal{N}(\mu_2, \sigma^2)$, так что $\mathbf{D}X = \mathbf{D}Y$. По двум независимым выборкам $X^{(n)} = (X_1, \ldots, X_n)$ и $Y^{(m)} = (Y_1, \ldots, Y_m)$ (возможно, разного объема) требуется проверить гипотезу *однородности* $H_0: \ \mu_1 = \mu_2$ при альтернативе $H_1: \ \mu_1 > \mu_2$. Типичный пример такой задачи – выявление эффекта нового метода лечения на группе из n пациентов посредством сравнения с контрольной группой из m пациентов, лечение которых проводится по старой методике.

Эта задача является для нас несколько новой, поскольку до сих пор мы имели дело только с одной выборкой. Тем не менее, она сводится к той, что мы только что рассмотрели в 3^0, с помощью следующих построений.

Рассмотрим сначала разность выборочных средних $\overline{X} - \overline{Y}$. Эта статистика имеет нормальное распределение со средним $\mathbf{E}(\overline{X} - \overline{Y}) = \mu_1 - \mu_2$ и дисперсией $\mathbf{D}(\overline{X} - \overline{Y}) = \mathbf{D}\overline{X} + \mathbf{D}\overline{Y} = \sigma^2(n^{-1} + m^{-1})$. Следовательно, при справедливости нулевой гипотезы $\mu_1 = \mu_2$ случайная величина

$$\xi = \frac{\overline{X} - \overline{Y}}{\sigma}\sqrt{\frac{nm}{n+m}}$$

имеет стандартное нормальное распределение $\mathcal{N}(0,\,1)$. Далее, нормированные выборочные дисперсии nS_X^2/σ^2 и mS_Y^2/σ^2 независимы и распределены по закону хи-квадрат с $n-1$ и $m-1$ степенями свободы соответственно. Так как для хи-квадрат распределения, как частного случая гамма-распределения, имеет место теорема сложения, то случайная величина $\eta = \left(nS_X^2 + mS_Y^2\right)/\sigma^2$ имеет хи-квадрат распределение с $n+m-2$ степенями свободы. Таким образом, мы приходим к двухвыборочной статистике Стьюдента

$$T_{n,m} = \frac{\xi}{\sqrt{\eta/(n+m-2)}} = \frac{\overline{X}-\overline{Y}}{\sqrt{nS_X^2+mS_Y^2}}\sqrt{\frac{nm(n+m-2)}{n+m}},$$

распределение которой при справедливости нулевой гипотезы есть распределение Стьюдента с $n+m-2$ степенями свободы.

Как и в случае одновыборочного критерия Стьюдента в 3^0 нетрудно показать, что при любых фиксированных C и σ функция мощности двухвыборочного критерия Стьюдента $T_{n,m} > C$ есть монотонно возрастающая функция параметра нецентральности $\Delta = (\mu_1-\mu_2)\sqrt{n+m-2}/\sigma$, так что критическая константа C определяется по заданному уровню значимости из уравнения $P(T_{n,m} > C) = 1 - S_{n+m-2}(C) = \alpha$ и равна квантили распределения Стьюдента: $C(\alpha) = S^{-1}_{n+m-2}(1-\alpha)$. Понятно, что при альтернативе $H_1:\ \mu_1 \neq \mu_2$ критическая константа $C(\alpha) = S^{-1}_{n+m-2}(1-\alpha/2)$.

При использовании этого критерия следует обратить особое внимание на предположение о равенстве дисперсий наблюдаемых случайных величин: $\sigma_X^2 = \sigma_Y^2$. Задача сравнения средних двух нормальных распределений с неравными дисперсиями и с гарантированным ограничением α на вероятность ошибки первого рода называется *проблемой Беренса–Фишера*. Известно лишь асимптотическое решение этой проблемы при больших n и m.

5^0. Сравнение дисперсий двух нормальных распределений при неизвестных средних (критерий Фишера). Независимые выборки $X^{(n)}$ и $Y^{(m)}$ берутся из соответствующих нормальных распределений $\mathcal{N}(\mu_1,\,\sigma_1^2)$ и $\mathcal{N}(\mu_2,\,\sigma_2^2)$, относительно параметров которого проверяется гипотеза $\sigma_1^2 = \sigma_2^2$ при альтернативе $\sigma_1^2 > \sigma_2^2$ с мешающими параметрами μ_1 и μ_2.

В этой задаче естественно рассмотреть критерий, основанный на статистике $F = nS_X^2/mS_Y^2$, которая распределена как

$$\frac{\chi^2_{n-1}}{\chi^2_{m-1}} \cdot \frac{\sigma_1^2}{\sigma_2^2}.$$

Функция мощности критерия $F > C$ (который называется *критерием Фишера* или *F-критерием*)

$$m\left(\frac{\sigma_1^2}{\sigma_2^2}\right) = P_{\mu_1,\mu_2,\sigma_1,\sigma_2}(F > C) = P\left(\frac{\chi^2_{n-1}}{\chi^2_{m-1}} > C \cdot \frac{\sigma_2^2}{\sigma_1^2}\right)$$

есть монотонно возрастающая функция отношения дисперсий σ_1^2/σ_2^2. Для ее вычисления необходимо знать распределение отношения двух независимых случайных величин, распределенных по закону хи-квадрат с $n-1$ и $m-1$ степенями свободы. Это так называемое *распределение Фишера* $F_{n-1,m-1}$, плотность которого

$$f_{n-1,m-1}(x) = \frac{\Gamma\left(\frac{n+m-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{m-1}{2}\right)} \cdot \frac{x^{\frac{n-1}{2}-1}}{(x+1)^{\frac{n+m-2}{2}}},\ x > 0,$$

вычисляется столь же просто, как это мы делали при выводе распределения Стьюдента. Таблицы распределения Фишера можно найти в ТМС. Критическая константа C критерия Фишера заданного размера α определяется как квантиль этого распределения: $C(\alpha) = F^{-1}_{n-1,m-1}(1-\alpha)$.

Мы завершим иллюстрацию методов построения критериев с помощью состоятельных оценок тестируемого параметра примером, в котором не всегда размер критерия совпадает с заданным уровнем значимости.

6^0. Проверка гипотезы о вероятности успеха в испытаниях Бернулли. Рассмотрим задачу проверки гипотезы $p = p_0$ против альтернативы $p < p_0$ о вероятности p успешного исхода в испытаниях Бернулли. Пример такой задачи – проверка гипотезы о вероятности наследования доминантного признака в опытах Менделя, когда альтернативная модель предписывает этой вероятности меньшее значение. Предлагаемый ниже метод решения позволяет строить критерии проверки такой гипотезы при альтернативах $p > p_0$ или $p \neq p_0$ посредством простой замены неравенства, определяющего критическую область, на обратное или двустороннее.

Итак, если мы располагаем выборкой $X^{(n)}$ из двухточечного распределения $\mathrm{B}(1, p)$, то относительная частота успешных испытаний (выборочное среднее) $\overline{X}$ является несмещенной оценкой p с минимальной дисперсией. В соответствии с предложенной выше идеологией проверки гипотез с помощью оценок тестируемого параметра мы должны отвергать гипотезу $p = p_0$ в пользу $p < p_0$, если $\overline{X} - p_0 < C$. Поскольку статистика $T = n\overline{X} = \sum_1^n X_k$ имеет биномиальное распределение $\mathrm{B}(n, p)$, а значение p_0 задано, то для вычисления функции мощности удобнее записать критическую область в виде $T < C$. Но статистика T принимает только целочисленные значения $0, 1, \ldots, n$, поэтому бессмысленно рассматривать дробные значения критических констант. Таким образом, мы приходим к наиболее удобной форме записи критической области в виде $T < C$, где C принимает значения $1, 2, \ldots, n$

Функция мощности такого критерия

$$m(p) = P_p(T \le C) = \sum_{k=0}^{C-1} p^k (1-p)^{n-k},$$

и поскольку проверяется простая гипотеза, то критическая константа C должна определяться по заданному уровню значимости α из неравенства

$$m(p_0) = \sum_{k=0}^{C-1} p_0^k (1-p_0)^{n-k} \le \alpha. \tag{2}$$

Очевидно, что чем больше C, тем больше мощность критерия, и поэтому $C(\alpha)$ следует выбирать как наибольшее целое число, удовлетворяющее неравенству (2). Размер критерия с таким $C(\alpha)$ не обязательно равен α, так что мы можем получить критерий уровня α, но не размера α (в предыдущих примерах с тестовыми статистиками, имеющими распределение непрерывного типа, мы имели критерии размера α). Более того, если p_0 настолько мало, что $(1-p_0)^n > \alpha$, то не существует таких C, при которых имеет место неравенство (2). В таком случае мы должны принимать нулевую гипотезу при любом результате статистического эксперимента, обеспечивая тем самым нулевой размер такого критерия “уровня α”.

При больших объемах выборки n можно использовать нормальные аппроксимации биномиального распределения, получая таким образом критерий, размер которого асимптотически ($n \to \infty$) равен α.

Статистика T асимптотически нормальна со средним np и дисперсией $np(1-p)$, поэтому неравенство (2) для определения критической константы имеет асимптотический аналог

$$\Phi\left(\frac{C - np_0}{\sqrt{np_0(1-p_0)}}\right) \le \alpha,$$

откуда $C(\alpha) \approx np_0 - \Phi^{-1}(1-\alpha)/\sqrt{np_0(1-p_0)}$. Легко понять, что такой метод построения критериев асимптотического уровня α применим для любой критической области, в задании которой используется асимптотически нормальная оценка тестируемого параметра (см. пояснения в предыдущем параграфе перед пунктом 5^0).

Этот пример показывает, что в случае дискретных распределений задача построения равномерно наиболее мощных критериев значительно усложняется, поскольку один из двух критериев одного и того же уровня α может иметь большую мощность только потому, что он имеет больший размер. Мы столкнемся с этой проблемой в следующем параграфе, но следует заметить, что современная теория наиболее мощных критериев обходит этот неприятный момент за счет расширения понятия статистического правила, вводя так называемые *рандомизированные* критерии. К сожалению, я не располагаю временем познакомить вас с этим замечательным объектом теории статистического вывода.

Мы закончим этот параграф, как и было обещано, формулировкой *принципа двойственности* между задачами проверки гипотез и доверительного оценивания. Пусть $A(\theta_0) \subset \mathfrak{X}^n$ – область принятия некоторого критерия уровня α, тестирующего гипотезу $H_0:\ \theta = \theta_0$, и пусть $A(\theta_0)$ определена при любом $\theta_0 \in \Theta$. Для каждого результата $x^{(n)}$ наблюдения случайной выборки $X^{(n)}$ введем подмножество $\Delta(x^{(n)})$ параметрического пространства Θ, положив $\Delta(x^{(n)}) = \{\theta:\ x^{(n)} \in A(\theta)\}$. *Тогда $\Delta(X^{(n)})$ есть $(1-\alpha)$-доверительное множество для параметра θ*, поскольку $P_\theta\left(\Delta(X^{(n)}) \ni \theta\right) = P_\theta\left(X^{(n)} \in A(\theta)\right) \ge 1-\alpha$.

Например, критерий Стьюдента проверки гипотезы $\mu = \mu_0$ о среднем значении нормального (μ, σ^2) распределения с неизвестной дисперсией σ^2 имеет область принятия (см. п. 3^0 данного параграфа)

$$A(\mu_0) = \frac{|\,\overline{X} - \mu_0\,|}{S}\sqrt{n-1} \le S_{n-1}^{-1}(1-\alpha/2).$$

Подставим в это неравенство вместо фиксированного μ_0 параметр μ и разрешим неравенство относительно μ. В результате получим доверительное утверждение (см. п. 4^0 предыдущего параграфа)

$$\overline{X} - St_\alpha/\sqrt{n-1} \leq \mu \leq \overline{X} + -St_\alpha/\sqrt{n-1},$$

в котором $t_\alpha = S_{n-1}^{-1}(1-\alpha/2)$.

Вы сами можете сопоставить доверительные интервалы, построенные в §6, с критериями из §7. При этом сопоставлении можно вывести полезное правило, касающееся доверительной оценки скалярного параметра θ. Если имеется состоятельный критерий проверки гипотезы $\theta = \theta_0$ при двусторонней альтернативе $\theta \neq \theta_0$, то его области принятия соответствует двусторонний доверительный интервал. Если же альтернативная гипотеза носит односторонний характер, то при альтернативе $\theta < \theta_0$ мы получаем верхнюю доверительную границу, а при $\theta > \theta_0$ – нижнюю.

Естественно, принцип двойственности применим и к доверительным интервалам, как статистическим правилам проверки гипотез: *гипотеза* $\theta \in \Theta_0$ *принимается тогда и только тогда, когда* $(1-\alpha)$-*доверительная область накрывает подмножество* Θ_0, и такое статистическое правило (критерий) гарантирует заданное ограничение α на вероятность ошибки первого рода.