

Построение доверительных интервалов с учётом априорной информации

А.В. Лохов, Ф.В. Ткачёв

Институт ядерных исследований РАН, Москва, 117312

A.V. Lokhov, F.V. Tkachov

Institute for Nuclear Research RAS, Moscow, 117312

Аннотация. Дается обзор методов решения фундаментальной задачи учёта априорной информации в виде одностороннего ограничения на оцениваемый параметр при построении доверительных интервалов.

Подробно рассматривается т.наз. метод предела чувствительности, дающий, как показывается, физически корректное решение задачи.

Строятся решения для ситуаций с непрерывным распределением в случае неотрицательного оцениваемого параметра, а также для дискретных распределений, в частности для пуассоновского процесса с фоном. Для этих же двух случаев построен наилучший верхний предел, учитывающий наличие априорной информации.

Для важного в физических приложениях случая пуассоновского распределения с неизвестным параметром распределения и известным фоном приведена таблица доверительных интервалов для параметра, а также представлена полная программа расчётов для произвольных доверительных уровней и значений фона (программа находится в открытом доступе).

Abstract. We review the methods of constructing confidence intervals that account for *a priori* information about a one-sided constraint on the parameter being estimated. We show that the method of sensitivity limit yields a correct solution of the problem. The solutions for the cases of a continuous distribution with non-negative estimated parameter and a discrete distribution, specifically a Poisson process with background, are derived. The best upper limit is constructed that accounts for the *a priori* information for the two cases. For the physically important case of Poisson distribution with unknown mean and a known background, a table of confidence intervals for the Poisson parameter is provided that correctly accounts for the known background, along with the software for calculating the confidence intervals for any confidence levels and magnitudes of the background (the software is freely available for download).

PACS: 29.85.Fj

1. Введение

Построение неймановских доверительных интервалов [1] для экспериментально измеряемых параметров — одна из важнейших задач обработки данных. При этом часто встречается случай, когда имеется априорная информация об оцениваемых параметрах, и важно корректно и полностью учесть эту информацию при построении интервалов.

Пример априорной информации — ограниченность области значений параметра, и проблема возникает в случае, когда экспериментальная оценка выпадает из этой области. Например, в эксперименте Троицк-ню-масс по измерению массы нейтрино в бета-распаде трития [2] параметр m_ν^2 заведомо больше либо равен нулю, однако полученное формальным фитированием значение m_ν^2 оказалось меньше нуля.

Другой важный случай, когда нужно учитывать априорную информацию, — построение доверительных интервалов в случае пуассоновского процесса при наличии пуассоновского же фона. Это обычная ситуация при изучении редких событий (в экспериментах по поиску двойного безнейтринного бета-распада [3], в экспериментах по осцилляциям нейтрино, например T2K, MINOS [4], и проч.).

Для решения этой задачи было предложено несколько решений. Решения можно разбить на две группы по способу использования свободы, заложенной в неймановской процедуре построения доверительных интервалов.

В первой группе решений априорная информация используется на стадии построения доверительных областей. К этому типу решений относится построение Фельдмана и Казинса [6], а также способ ограничения интервалов с помощью функции мощности, предложенный Кауэном и др. (т.наз. метод CCGV [7]). Однако интервалы, построенные, например, по рецепту Фельдмана и Казинса (способ, имеющий огромную цитируемость), не позволяют произвести объективное сравнение результатов разных экспериментов (эксперимент с заведомо худшей чувствительностью и систематикой может дать лучший интервал).

Другая группа решений предполагает использование априорной информации о параметрах при выборе самой оценки (estimator). К этой группе, по сути, принадлежит конструкция доверительных интервалов, указанная в работе [8], где, однако, рассматривается только случай оценивания по методу максимального правдоподобия, рассуждения не вполне прозрачны (и местами искусственны), и даже не приводится явная формула для оценки. По этим причинам не видно, что окончательный рецепт может быть использован с любыми методами оценивания. Надо полагать, что именно поэтому

конструкция доверительных интервалов [8], физически, как выясняется, корректная, оставалась, судя по истории цитирования, практически неизвестной.

Полное решение, не зависящее от способа оценивания (будь то метод моментов, метод наименьших квадратов или метод максимального правдоподобия), с явной формулой для оценки было найдено в работе [9], где, в отличие от работы [8], дана ясная графическая и аналитическая интерпретация для всей конструкции.

Оказывается, что эта конструкция для нефизических, т.е. выпадающих за априорную границу значений оценки совпадает с эмпирическим рецептом т.наз. предела чувствительности (использованным, например, среди прочих методов для представления результатов эксперимента по измерению массы нейтрино в Майнце [10]), поэтому мы будем называть рассматриваемую конструкцию методом предела чувствительности.

Возможно, главным достоинством метода предела чувствительности является тот факт, что он позволяет напрямую объективно сравнивать результаты различных экспериментов, не требуя повторной обработки. Именно так, например, новые результаты эксперимента Троицк-ню-масс [2] сравниваются как с упомянутым старым измерением в Майнце, так и с результатом старой обработки данных Троицк-ню-масс [11].

В этом отношении метод качественно превосходит построение Фельдмана и Казинса [6], где в нефизической области доверительный интервал зависит от экспериментального значения оценки параметра и даже противоестественно уменьшается по мере удаления от физической границы в нефизическую область. Фактически рецепт Фельдмана и Казинса не даёт результатов, которые можно было бы сравнивать.

Интересно, что применение второго подхода (т.е. учёт априорной информации ещё при построении оценки) позволяет также указывать корректные и оптимальные односторонние (сверху или снизу) ограничения на параметры [12].

Для распространения решения на случай дискретных распределений также проще всего (и понятней) использовать процедуры работ [9], [12].

Сначала в разделе 2 мы напоминаем неймановскую конструкцию, формулируя ее в виде, удобном для вывода метода предела чувствительности. При этом особо рассматривается дискретный случай, где — в силу самой дискретности — необходимо отказаться от равенств, описывающих вероятностное содержание доверительных интервалов, в пользу неравенств. Обсуждается построение интервалов симметричных, несимметричных, а также способ построения Стерна, Кроу и Гарднера [5].

В разделах 3-5 дается обзор конструкций Кауэна и др., Фельдмана и Казинса и Манделькерна и Шульца соответственно.

В разделе 6, следуя [9], проводится построение метода предела чувствительности для непрерывного случая, когда априорная информация об измеряемом параметре имеет вид неравенства $\theta \geq 0$. В разделе 7 рассматривается дискретный случай пуассоновского процесса с неизвестным μ , но известным значением пуассоновского фона b . Наконец, в разделе 8 строятся наилучшие верхние пределы для непрерывного и дискретного случая. Выводы суммированы в разделе 9.

Подчеркнем, что в данной работе рассматривается только неймановская процедура построения доверительных интервалов. Такие интервалы имеют ясную интерпретацию в рамках частотного подхода к интервальному оцениванию. Другой используемый подход — байесовский — хотя и может быть проинтерпретирован в терминах статистических ансамблей [13], но применяемая экспериментаторами процедура построения байесовских интервалов использует неизвестную априорную функцию плотности вероятности для параметра, что нарушает условия применимости теоремы Байеса в узком смысле и затрудняет интерпретацию результатов. Мы не рассматриваем в настоящей работе и построение доверительных интервалов методом CLs [14], поскольку он не имеет ясной интерпретации ни в рамках частотного, ни байесовского подхода, хотя и широко используется, например, в представлении результатов поиска бозона Хиггса [15].

2. Неймановское построение

а) Непрерывный случай

Приведем вначале стандартную процедуру построения доверительных интервалов по Нейману [1] и зафиксируем основные обозначения, которые будут использоваться в дальнейшем.

Пусть $\hat{\theta}$ — обычная оценка для неизвестного параметра θ , то есть оценка, полученная без учёта априорного ограничения (например, фундаментальным методом моментов [16], [17]).

Случайная величина $\hat{\theta}$ является функцией набора экспериментальных данных X : $\hat{\theta} = \hat{\theta}(X)$. Плотность вероятности этой величины $d_{\theta}(\hat{\theta})$ зависит от θ как от параметра. Предполагается, что плотность вероятности известная, несингулярная функция, как того требует стандартная неймановская процедура построения доверительных интервалов [1]. Плотность $d_{\theta}(\hat{\theta})$ содержит в себе всю информацию об измерении параметра θ в эксперименте, включая метод оценивания этого параметра.

Пусть α, α' — малые неотрицательные величины. Определим $L_{\alpha}(\theta)$ и $U_{\alpha'}(\theta)$ как

$$P(-\infty < \hat{\theta} < L_{\alpha}(\theta)) = \alpha, \quad P(U_{\alpha'}(\theta) < \hat{\theta} < +\infty) = \alpha' \quad (1)$$

Вероятность получить величину оценки меньше $L_{\alpha}(\theta)$ равна α , а больше $U_{\alpha'}(\theta)$ — α' .

Таким образом, $L_{\alpha}(\theta)$ соответствует определённой в разделе 9.1.1 [16] величине Z_{α}

Если предположить также, что $L_{\alpha}(\theta)$ и $U_{\alpha'}(\theta)$ обратимые функции θ , то (1) можно переписать в виде

$$P(l_{\alpha}(\hat{\theta}) < \theta) = \alpha, \quad P(\theta < u_{\alpha'}(\hat{\theta})) = \alpha' \quad (2)$$

где $u_{\alpha} = U_{\alpha}^{-1}$, $l_{\alpha} = L_{\alpha}^{-1}$. Из (2) следует, что вероятность получить случайную величину $l_{\alpha}(\hat{\theta})$ ($u_{\alpha'}(\hat{\theta})$) меньшую (большую), чем неизвестное истинное значение θ равна α (α').

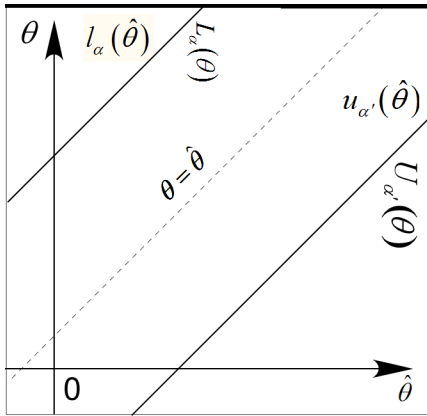


Рис. 1. Показаны функции $\theta = l_{\alpha}(\hat{\theta})$ и $\theta = u_{\alpha'}(\hat{\theta})$ (или, с другой точки зрения, $\hat{\theta} = L_{\alpha}(\theta)$ и $\hat{\theta} = U_{\alpha'}(\theta)$). В общем случае диагональ $\theta = \hat{\theta}$ не лежит между двумя кривыми (на последующих иллюстрациях диагональ отображаться не будет). Две сплошные линии будут использоваться в дальнейшем для случая, когда они составляют стандартную симметричную систему доверительных интервалов с уровнем доверия $\beta = 1 - 2\alpha$. Чем меньше величина β , тем уже доверительные интервалы.

Можно переписать (1) в виде

$$P(L_{\alpha}(\theta) < \hat{\theta} < U_{\alpha'}(\theta)) = 1 - \alpha - \alpha' \equiv \beta \quad (3)$$

Тогда из эквивалентного выражения

$$P(u_{\alpha'}(\hat{\theta}) < \theta < l_{\alpha}(\hat{\theta})) = \beta \quad (4)$$

следует, что с вероятностью β (доверительный уровень, например, 90%) случайный интервал $[u_{\alpha'}(\hat{\theta}), l_{\alpha'}(\hat{\theta})]$ накрывает неизвестное значение θ .

Выбирая $\alpha = \alpha' = (1 - \beta)/2$, получим стандартный симметричный доверительный пояс (систему доверительных интервалов).

Такое построение содержит следующую свободу. Зафиксируем уровень доверия β (например, $\beta = 90\%$, далее везде полагаем β фиксированным). Выберем две функции L , U , удовлетворяющие условию

$$P(L(\theta) < \hat{\theta} < U(\theta)) = \beta \quad (5)$$

Если эти функции также монотонны, то существуют обратные им $u = U^{-1}$, $l = L^{-1}$.

Тогда из эквивалентного выражения

$$P(u(\hat{\theta}) < \theta < l(\hat{\theta})) = \beta \quad (6)$$

снова следует, что случайный интервал $[u(\hat{\theta}), l(\hat{\theta})]$ накрывает неизвестное значение θ с вероятностью β .

Заметим, что линия $\theta = u(\hat{\theta})$ расположена ниже $\theta = u_{1-\beta}(\hat{\theta})$ (то есть $u(\hat{\theta}) \leq u_{1-\beta}(\hat{\theta})$).

Аналогично, линия $\theta = l(\hat{\theta})$ ограничена снизу. Каждая пара таких линий задаёт допустимый доверительный пояс для доверительного уровня β .

На Рис. 2 в дополнение к симметричному доверительному поясу с уровнем доверия β , введен также более узкий пояс для меньшего уровня доверия $\tilde{\beta} = 1 - 2(1 - \beta) = 1 - 4\alpha < \beta$ (наклонные пунктирные линии). Обозначения различных точек пересечения и горизонтальных линий введены для удобства: одинаково обозначенные на последующих иллюстрациях точки совпадают точкам на данном рисунке.

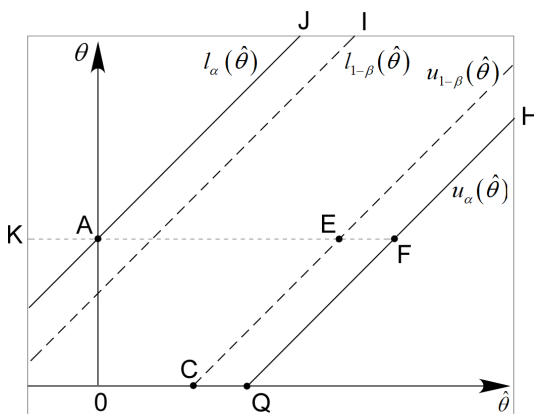


Рис. 2. Пары сплошных и пунктирных наклонных линий ограничивают симметричные доверительные пояса для доверительных уровней $\beta = 1 - 2\alpha$ и

$\tilde{\beta} = 1 - 2(1 - \beta) = 1 - 4\alpha$; см. Рис. 1. На рисунке указаны функции, соответствующие линиям. Точка А - пересечение вертикальной оси с линией $\theta = l_\alpha(\hat{\theta})$. Точка А задаёт горизонтальную линию KF с другими точками пересечений. Точки С и Q - пересечения линий $\theta = u_{1-\beta}(\hat{\theta})$ и $\theta = u_\alpha(\hat{\theta})$ с горизонтальными осями.

Обозначим ординату точки А (пересечение вертикальной оси и прямой KF) θ_A :

$$\theta_A = l_\alpha(0) \quad (7)$$

Числа $\theta_C < \theta_E < \theta_F$ суть ординаты точек С, Е, F:

$$\theta_C = U_{1-\beta}(0), \quad \theta_E = U_{1-\beta}(\theta_A), \quad \theta_F = U_\alpha(\theta_A) \quad (8)$$

б) Дискретный случай

Будем рассматривать ситуацию, когда в эксперименте детектируется число событий n . Число зарегистрированных событий может быть распределено, например, по Пуассону:

$$P_\mu(n) = \frac{\mu^n}{n!} e^{-\mu} \quad (9)$$

Здесь μ - параметр пуассоновского распределения, среднее число событий. (Дальнейшие рассуждения не зависят от конкретного вида распределения.)

Дискретность распределения приводит к некоторым изменениям в построении доверительных интервалов по сравнению с непрерывными распределениями.

Для построения доверительных интервалов выберем, как обычно, доверительный уровень α (например, 95%). По определению необходимо указать такие интервалы, которые будут содержать неизвестное истинное значение μ_0 в доле экспериментов равной α :

$$P(\mu_0 \in [\mu_1, \mu_2]) = \alpha \quad (10)$$

Вследствие дискретности распределения необходимо ослабить условие (10), изменив его на неравенство " \geq " [6].

Для каждого значения μ можем указать значения n , удовлетворяющие условию:

$$P_\mu(n \in [n_1(\mu), n_2(\mu)]) \geq \alpha \quad (11)$$

Рассмотрим сначала односторонний интервал:

$$P(n \leq n_\alpha(\mu)) \geq \alpha \quad (12)$$

Для определения доверительного интервала для μ можно сделать преобразование выражения в скобках:

$$P(\mu \geq \mu^*(n)) \geq \alpha \quad (13)$$

Неравенство (13) означает, что в доле $\geq \alpha$ экспериментов (измерений числа n) будет получено значение n такое, что неизвестное истинное значение параметра удовлетворяет условию $\mu \geq \mu^*(n)$.

В случае дискретного распределения для величины n функция $\mu^*(n)$ определена, на первый взгляд, не однозначно (каждому значению n соответствует набор значений μ^* , при которых выполнено неравенство (13)). Ниже мы покажем, что условие консервативности интервала [6] (требование выполнения (13) для любых фиксированных μ) определяет единственное значение μ^* для каждого n . Это значение μ^* и будет границей доверительного интервала.

Для нахождения $\mu^*(n)$ введём следующее обозначение: пусть μ_n - значение параметра, при котором выполняется точное равенство $P(n \leq n_\alpha(\mu_n) - 1) = \alpha$ (при этом $P(n \leq n_\alpha(\mu_n + \varepsilon) - 1) < \alpha$ и $P(n \leq n_\alpha(\mu_n + \varepsilon)) > \alpha$), где ε - сколь угодно малое положительное число.

Покажем, что если выбрать в качестве $\mu^*(n)$ величину $\mu^*(n) = \mu_n + \varepsilon$, то условие консервативности не будет выполнено.

При таком выборе $\mu^*(n)$, если истинное значение μ лежит в интервале $(\mu_n, \mu_n + \varepsilon)$, условие (13) будет нарушаться. Действительно, в таком случае μ окажется больше либо равной случайной величине $\mu^*(n) = \mu_n + \varepsilon$ только в тех случаях, когда при измерении будут получены значения $n = n_\alpha(\mu) - 1$, $n = n_\alpha(\mu) - 2$ и т.д. Вероятность получить такие значения n равна $P(n \leq n_\alpha(\mu) - 1) < \alpha$ (согласно выбору величины μ_n). Следовательно, интервал $[\mu_n + \varepsilon, +\infty)$ по определению не является доверительным интервалом.

При выборе в качестве $\mu^*(n)$ случайной величины μ_n , условие $\mu \geq \mu_n$ будет выполнено в доле измерений $\geq \alpha$, что удовлетворяет определению доверительного интервала для параметра μ .

Таким образом, односторонний доверительный интервал в случае дискретного распределения задаётся формулой $\mu \geq \mu_n$, где величина μ_n определяется условием:

$$P(n \leq n_\alpha(\mu_n) - 1) = \alpha.$$

Аналогично можно рассмотреть интервал $P(n \geq n_\alpha(\mu)) \geq \alpha$. Соответствующая верхняя граница доверительного интервала для μ будет определяться формулой $\mu \leq \mu'_n$. Здесь μ'_n задаётся условием $P(n \geq n_\alpha(\mu'_n) + 1) = \alpha$.

Дискретность распределения вносит дополнительную неопределённость в построение двусторонних доверительных интервалов. Так, например, можно задать двусторонний доверительный интервал, объединяя верхнюю и нижнюю границы односторонних интервалов:

$$P(n_1(\mu) \leq n \leq n_2(\mu)) \geq \alpha \Rightarrow P(\mu_1(n) \leq \mu \leq \mu_2(n)) \geq \alpha \quad (14)$$

В качестве μ_1 и μ_2 выбираются соответствующие значения μ_n и μ'_n . Значения этих величин для каждого n , очевидно, зависят от доверительного уровня α .

Такой выбор приводит к приблизительно симметричному доверительному интервалу. Приблизительность состоит в том, что вероятностное содержание областей $n < n_1(\mu)$ и $n_2(\mu) < n$ может не быть одинаковым.

Для указания интервала можно также руководствоваться физическими соображениями. Так, если важно получить ограничение на параметр сверху, то можно потребовать выполнения условия $P(n < n_1(\mu)) < (1 - \alpha)/2$. В таком случае вероятностное содержание области выше верхней границы будет $< (1 - \alpha)/2$. Построим нижнюю границу доверительного интервала, используя значения $n_1(\mu)$ как левую границу в условии $P(n_1(\mu) \leq n \leq n_2(\mu)) \geq \alpha$. В этом случае может оказаться, что условие $P(n > n_2(\mu)) < (1 - \alpha)/2$ не выполнено. Нижняя граница доверительного интервала при таком построении может не совпадать с границей одностороннего интервала, полученной из условия (12). В то же время такой интервал будет менее перекрывающим, то есть вероятностное содержание соответствующей доверительной области будет ближе к заданному значению α (как справедливо указано в [6], это качество — перекрывание — является нежелательным для доверительных интервалов).

Аналогично можно указать интервал, если важно более строгое ограничение на измеряемый параметр снизу.

Ещё один способ построения доверительных интервалов предложен Стерном, Кроу и Гарднером [5]. Идея способа заключена в конструировании доверительной области добавлением в неё точек в порядке убывания соответствующей вероятности (в отличие от построения симметричной системы интервалов). Впервые способ был применен к биномиальному (дискретному) распределению, хотя возможно его распространение и на непрерывные распределения. Такое построение приводит в общем случае к несимметричным интервалам, но соответствующие интервалы имеют меньшую длину. Как показал Кроу, такой способ даёт доверительную полосу с наименьшей общей площадью.

Таким образом, для каждого измерения величины n можно указать доверительный интервал для неизвестного параметра μ в форме

$$\mu \in [\mu_{n\alpha}, \mu'_{n\alpha}]$$

несколькими способами. Выбор конкретного способа задания доверительного интервала, как обычно в задачах обработки данных, остаётся за экспериментатором.

3. Построение CCGV

Перейдём к основным, известным из литературы, способам учёта априорной информации в процедуре построения доверительных интервалов. Рассмотрим вначале попытки учёта априорной информации при построении доверительной области (решения первой группы).

Кауэн и др. предложили метод, получивший название ограничения интервалов с помощью функции мощности [7]. Этот метод состоит в следующем. Выбирается некий статистический критерий со статистикой $q_\mu = q_\mu(x)$ (для удобства выбирается статистика, увеличивающаяся при увеличении несогласия между экспериментальными данными x и величиной параметра μ). Для этого критерия строится функция мощности, которая и является ключевым элементом этого построения:

$$M_{\mu'}(\mu) = P(q_\mu(x) > q_{\mu, \text{crit}} | \mu')$$

Рассматриваются две гипотезы: $\mu' = 0$ (сигнал отсутствует) и $\mu > 0$ (существует ненулевой сигнал) и соответствующая функция мощности $M_0(\mu)$. Выбирается некоторое пороговое значение функции мощности M_{\min} , и область значений μ разбивается на две: если для некоторого значения μ значение $M_0(\mu)$ оказывается ниже порогового, то считается, что чувствительность к этому параметру слишком мала, и такие значения μ не

могут быть проверены. Таким образом, μ не включается в доверительный интервал для некоторого конкретного набора экспериментальных данных, если 1) μ отвергается критерием q_μ при заданном доверительном уровне α , и 2) чувствительность к μ достаточна, т.е. $M_0(\mu) \geq M_{\min}$.

Все значения μ , не удовлетворяющие условиям 1) или 2) составляют искомый доверительный интервал.

Вероятность, с которой доверительный интервал накрывает данное значение μ , равна 100%, для тех μ , для которых функция мощности меньше порогового значения, и α для тех μ , для которых мощность больше либо равна пороговой. Выбор порогового значения M_{\min} остаётся за экспериментатором.

Таким образом, построение Кауэна и др.

1) содержит перекрывание (избыточное вероятностное содержание доверительной области) в самой процедуре;

2) не даёт ясной интерпретации значения M_{\min} , которое и определяет конечный вид интервалов;

3) не решает проблему уменьшения доверительных интервалов в нефизической области.

4. Построение Фельдмана и Казинса

Построение доверительных интервалов, предложенное Фельдманом и Казинсом [6], так же как и способ Стерна, Кроу и Гарднера, основано на специальном порядке добавления точек в доверительную область. Порядок добавления точек определяется отношением правдоподобий.

Рассмотрим следующий пример. Пусть в эксперименте измеряется число событий n , распределённое по Пуассону, с параметром $(\mu + b)$. Здесь μ - неизвестный параметр, который необходимо оценить, b - известный фон. Если обозначить вероятность получить в эксперименте число событий n при $\mu = \mu_1$ через $P(n | \mu_1)$, то вышеупомянутое отношение правдоподобий есть

$$R(n) = \frac{P(n | \mu_1)}{P(n | \mu_{best})} \quad (15)$$

где величина $\mu_{best} = \max(0, n - b) - 1$ максимизирует $P(n | \mu)$ при данном n ; 2)

неотрицательна. Далее, для каждого значения μ точки n добавляются в доверительную область в порядке уменьшения соответствующей величины $R(n)$ до тех пор, пока суммарная вероятность, содержащаяся в доверительной области, не достигнет заданного доверительного уровня.

Такой рецепт построения доверительных интервалов, как отмечено в самой работе [6], не решает важную проблему, возникающую в случае меньшего, чем фон, числа событий: доверительный интервал оказывается тем меньше, чем меньшее число событий получено в результате эксперимента. Следовательно, независимо от величины фона в эксперименте может получиться сколь угодно сильное ограничение на величину сигнала.

Аналогичная проблема возникает при применении рецепта [6] и для ограниченных параметров непрерывных распределений. Чем дальше в нефизической области оказывается экспериментальное значение оценки, тем меньший доверительный интервал такому значению соответствует. Рецепт Фельдмана и Казинса даёт в этом случае парадоксальный результат — наиболее ненадёжные результаты (оценки, лежащие в нефизической области) дают наиболее сильное ограничение на оцениваемый параметр.

Обе указанных проблемы построения [6] приводят к тому, что невозможно провести сравнение не только результатов разных экспериментов, но и результатов одного эксперимента (например, двух разных сеансов), если результаты представлены в форме доверительных интервалов, полученных по рецепту Фельдмана и Казинса.

Невозможность сравнивать доверительные интервалы характерна для всех решений из первой группы (см., например, попытку дальнейшей модификации самих интервалов и порядка построения доверительной области в работе [18]).

Этого существенного недостатка лишены решения второй группы.

5. Конструкция Манделькерна и Шульца

Предыдущие исследователи использовали изменение порядка добавления точек в доверительную область для получения доверительных интервалов с учётом дополнительной информации о параметрах.

Однако их построения приводят к нефизически коротким доверительным интервалам вблизи границы физически допустимых значений параметра. Это происходит оттого, что способ построения оценки фактически не учитывает физического ограничения на

интересующие параметры, и используется та же оценка, что и в случае без априорных ограничений.

Манделькерн и Шульц [8] предложили использовать в решении задачи другую оценку. В частном случае, рассматриваемом в работе [8], подходящая оценка находится с использованием метода максимального правдоподобия.

В функцию правдоподобия вводится фактор – функция Хевисайда, явным образом отражающая условие ограничения параметра, и получается оценка, всегда лежащая в физической области для данной задачи. Между тем, введение такого фактора выглядит искусственно (постулируется), а корректность рецепта становится очевидной лишь постфактум — сравнением распределения для оценки из [8] с общим решением [9].

Дальнейший вывод доверительных интервалов производится без каких-либо дополнительных предположений.

Отметим, что такая процедура не является байесовской: исключение нефизических значений параметра не является аналогом введения равномерной байесовской априорной функции распределения.

Частное решение [8], вообще говоря, корректно (хотя, повторим, убедиться в этом проще всего, сравнив его с общим решением): такое построение

1) в отличие от построения Фельдмана и Казинса формально решает проблему меньшего чем фон измеренного числа событий (для пуассоновского процесса с фоном) и уменьшения доверительного интервала для отрицательных значений оценки заведомо положительного параметра гауссовского распределения;

2) даёт правильное вероятностное содержание доверительной области, в отличие от flip-flop рецептов, нарушающих условия (3) и (11).

С другой стороны, рецепт построения оценки здесь основан на методе максимального правдоподобия, и исключает целый класс задач, где изначально выбран другой метод оценивания.

Вероятно, все эти причины и обусловили тот факт, что подход Манделькерна и Шульца почти не использовался на практике.

6. Обоснование корректной системы доверительных интервалов для непрерывного случая

Полное решение задачи о доверительных интервалах для параметра непрерывного распределения с учётом априорной информации об ограниченной области значений параметра даётся в работе [9].

Как уже отмечалось, ключевым элементом в построении доверительных интервалов является оценка (estimator). Можно поставить вопрос: как правильно выбрать оценку $\hat{\theta} = \hat{\theta}(X)$, если заранее известно, что $\theta \geq 0$?

Определяющим свойством любой оценки является то, что она даёт значение наиболее близкое к неизвестному θ . Тогда определим новую оценку как:

$$\tilde{\theta} = \max(\hat{\theta}, 0) \quad (16)$$

Очевидно, что $\tilde{\theta}$ даёт оценки, которые заведомо ближе к неизвестному значению θ , чем $\hat{\theta}$. Такая оценка содержит как статистическую информацию, заключенную в обычной оценке $\hat{\theta}$, так и априорную информацию о том, что $\theta \geq 0$. После этого остаётся построить доверительные интервалы для новой оценки $\tilde{\theta}$.

Прежде чем продолжать построение, обсудим определение (16).

Распределение вероятности для $\tilde{\theta}$ имеет следующий вид:

$$\tilde{d}_{\theta}(\tilde{\theta}) = H(\tilde{\theta}) d_{\theta}(\tilde{\theta}) + c_{\theta} \delta(\tilde{\theta}) \quad (17)$$

где $H(t)$ - обычная функция Хевисайда, $\delta(t)$ дираковская δ -функция и

$$c_{\theta} = \int_{-\infty}^0 d\hat{\theta} d_{\theta}(\hat{\theta}) \quad (18)$$

Следовательно, возникает дополнительная сложность, связанная с сингулярным вкладом в (17). Для работы с таким вкладом можно использовать стандартные методы, либо применить специальный приём. Применение стандартного подхода (регуляризации) описано в работе [9].

Прейдем к описанию специального приёма. Воспользуемся следующим наблюдением. Определение (16) означает, что случайные значения обычной оценки $\hat{\theta}$ в итоге переносятся в точку ноль и скапливаются в этой точке. Это означает, что все такие значения становятся неразличимыми: они дадут нулевое значение модифицированной оценки $\tilde{\theta}$ — и, следовательно, один и тот же доверительный интервал. Тогда все неположительные значения $\hat{\theta}$ дадут в итоге один и тот же доверительный интервал $[0, \text{const}]$, где константа не зависит от $\hat{\theta}$.

Как только это стало понятным, построение доверительных интервалов может быть произведено целиком в терминах обычной оценки $\hat{\theta}$; модифицированная оценка $\tilde{\theta}$ используется лишь как дополнительное условие: итоговые системы доверительных интервалов должны удовлетворять условию, что все значения $\hat{\theta}$ за априорной границей должны давать одинаковый доверительный интервал.

Это дополнительное условие имеет ясный экспериментальный смысл: условие можно сформулировать как требование устойчивости доверительных интервалов к возможным экспериментальным дефектам, что придаёт всей конструкции дополнительный физический смысл. Следует, однако, помнить, что приведённое рассуждение прозрачно и конкретно и само по себе не требует неких метафизических обоснований: построение интервалов в терминах обычной оценки $\hat{\theta}$ с вышеупомянутым дополнительным условием эквивалентно непосредственному построению доверительных поясов для модифицированной оценки (16), которая включает в себя априорную информацию наиболее простым и прозрачным образом.

Горизонтальные деформации. Приём горизонтальных деформаций основан на следующих свойствах доверительных поясов для фиксированного β . Если для данного θ величина $U(\theta)$, определённая формулой (5), смещается к её нижней границе $U_{1-\beta}(\theta)$, тогда соответствующее значение величины $L(\theta) \rightarrow -\infty$ (аналогично можно смещать L к её верхней границе). Если L и U могут быть таким образом деформированы (при условии постоянного сохранения непрерывности и монотонности), то существуют корректно определённые обратные функции $u = U^{-1}, l = L^{-1}$ на каждом шаге деформаций — то есть существует доверительный пояс для уровня доверия β .

На наших графиках такие деформации происходят в горизонтальных направлениях. Это показано на Рис. 3: жирные кривые должны лежать вне внутреннего пояса и могут только приближаться к границам этого пояса (в горизонтальном направлении, как показано на Рис. 3 стрелками) с одной стороны, при этом с другой стороны кривая удаляется от границы на бесконечность.

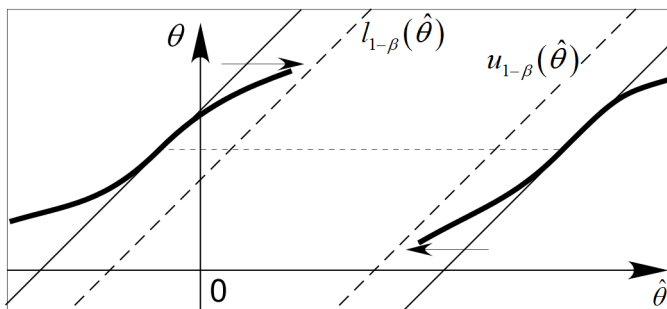


Рис. 3. Пары сплошных и пунктирных наклонных прямых ограничивают симметричные доверительные пояса для доверительных уровней $\beta = 1 - 2\alpha$ и $\tilde{\beta} = 1 - 2(1 - \beta) = 1 - 4\alpha$ (ср. Рис. 1). Две жирных кривых показывают возможный выбор l , u для доверительного уровня $\beta = 1 - 2\alpha$. Стрелки указывают направление горизонтальных деформаций, обсуждаемых в тексте.

Как только одна из жирных кривых пересекает границу симметричного доверительного пояса (сплошные наклонные прямые), вторая жирная кривая пересекает другую границу, как это показано с помощью горизонтальной пунктирной линией.

Описанная свобода была использована в работе [6], где пара L, U была выбрана на основе веры в "магию правдоподобий". Выбор системы интервалов из работы [6] показан на Рис. 4.

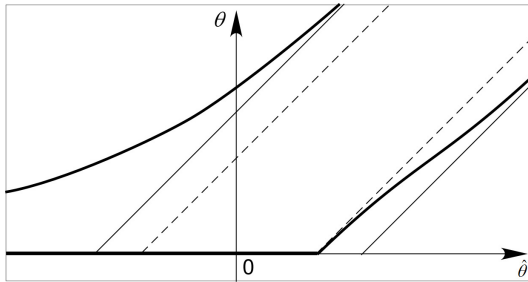


Рис. 4. Доверительный пояс (ограниченный жирными линиями) определённый в работе [6] (ср. с Рис. 10 из указанной работы). Остальные линии те же что и на Рис. 3. Правой жирной кривой, приближающейся к пунктирной границе, соответствует левая кривая, уходящая на бесконечность с приближением к горизонтальной оси, ср. Рис. 3.

Наконец, можно перейти к пределу, деформируя L таким образом, что её часть становится частью кривой $L_{1-\beta}(\theta)$ (см. Рис. 5). Это может привести к тому, что L (и U) перестанут быть непрерывными на границе этой части. Однако если соответствующие обратные функции $u = U^{-1}, l = L^{-1}$ непрерывно приближаются в пределе к хорошо определённым непрерывным монотонным (неубывающим) функциям величины $\hat{\theta}$, то система доверительных интервалов (6) будет непрерывно приближаться к хорошо определённому результату, а предельный доверительный пояс будет настолько же хорош, как и любой допустимый доверительный пояс для оценки параметров.

Корректное построение доверительных интервалов можно провести с использованием описанного приема горизонтальных деформаций. В обозначениях Рис. 2 и для фиксированного доверительного уровня β рассмотрим доверительный пояс соответствующий двум функциям l, u , выбранным так, как показано на Рис. 5 (ср. также Рис. 3 и 4).

2) симметричный пояс для уровня доверия $\tilde{\beta} = 1 - 2(1 - \beta) = 1 - 4\alpha$, имеющий вид $[u_{1-\beta}(\hat{\theta}), l_{1-\beta}(\hat{\theta})]$.

Верхняя граница доверительного интервала $[0, \text{const}]$, отвечающего всем нефизическим значениям $\hat{\theta}$, и определяет так называемый *предел чувствительности*. Величина предела чувствительности, очевидно, не зависит от конкретной величины $\hat{\theta}$, а задаётся только ошибкой определения $\hat{\theta}$. Таким образом, предел чувствительности отражает величину экспериментальной ошибки и даёт объективное представление результатов эксперимента, если экспериментальная оценка параметра лежит в нефизической области.

7. Случай дискретных распределений

Рассмотрим снова случай, когда в эксперименте детектируется число событий n , причём n имеет распределение Пуассона (9) с параметром μ . Учтём теперь наличие фоновых событий. Число фоновых событий - измеряемая величина, следовательно, в общем случае известно распределение $P_{\beta}(b)$. Здесь β - неизвестное истинное значение для среднего числа фоновых событий.

В работе [6] рассматривался случай, когда среднее число фоновых событий известно точно и равно b . Тогда число событий, зарегистрированных в эксперименте, будет распределено по Пуассону со средним $(\mu + b)$:

$$P_{\mu}(n) = \frac{(\mu + b)^n}{n!} e^{-(\mu + b)} \quad (19)$$

Используем теперь дополнительную информацию о фоне и построим доверительные интервалы для параметра μ в духе рассуждений [9]. Основная идея состоит в том, что априорная информация о параметре может быть снова учтена в выборе оценки (estimator) для данного параметра. После этого построение интервалов производится автоматически (ср. с предыдущими попытками [6] модификации непосредственно доверительных интервалов, содержащими в себе элемент произвола). В частном случае, когда среднее значение для фона известно точно, можно применить прием из [9] и выбрать в качестве оценки (estimator) следующую величину:

$$\widetilde{(\mu + b)} = \max(n, b) \quad (20)$$

Очевидно, что измеряемая в эксперименте величина n может быть выбрана в качестве оценки для $(\mu + b)$, но, в отличие от (20), допускает отрицательные значения для μ . Измеряя n и используя оценку (20), в результате вычитания постоянного фона b получим заведомо неотрицательную оценку параметра μ .

Распределение для оценки (20) позволяет строить доверительные интервалы для $(\mu + b)$ и, следовательно, для величины μ . Дискретность распределения для величины n (19) позволяет провести следующее рассуждение. Для любого заданного μ известна вероятность получить при измерении значение $n \leq b$, она равна:

$$P(n \leq b) = \sum_{n=0}^{[b]} \frac{(\mu + b)^n}{n!} e^{-(\mu + b)} \quad (21)$$

где $[b]$ обозначает, как обычно, целую часть числа. Тогда, используя оценку (20), в результате измерений будем получать значение этой оценки равное b с вероятностью (21).

Таким образом, распределение вероятностей для случайной величины $\max(n, b)$ будет состоять из распределения (19) при значениях $n > b$, и вероятности (21) получить при измерениях величину b (Рис. 7).

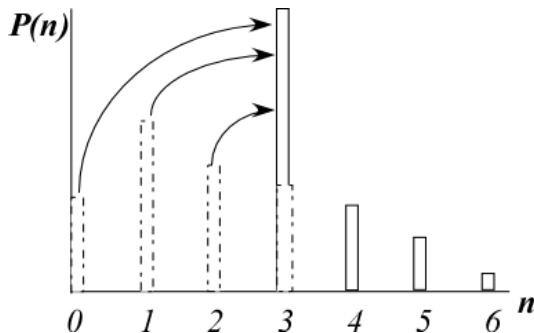


Рис. 7. Схема изменения распределения вероятностей при переходе от оценки $\widetilde{(\mu + b)} = n$ (штрихованный график в области $n \leq b$) к оценке вида $\widetilde{(\mu + b)} = \max(n, b)$ (сплошной график) при величине фона $b = 3$. При $n > b$ значения вероятностей для обычной и модифицированной оценок совпадают.

Используя полученное распределение вероятностей можно, следуя предыдущему разделу, построить доверительный интервал для величины $(\mu + b)$ и, соответственно, интересующего нас параметра μ .

Для практического применения удобно указать доверительные интервалы в переменных μ и n (Рис. 8). В таком представлении при получении в эксперименте значения n_0 можно непосредственно указать доверительный интервал для параметра μ .

Заметим, что все значения $n \leq b$ неразличимы после перехода к оценке вида (20). Поэтому всем этим значениям соответствует один и тот же доверительный интервал. Верхняя граница такого интервала соответствует так называемому *пределу чувствительности*. Она несет информацию о величине фона в конкретном эксперименте.

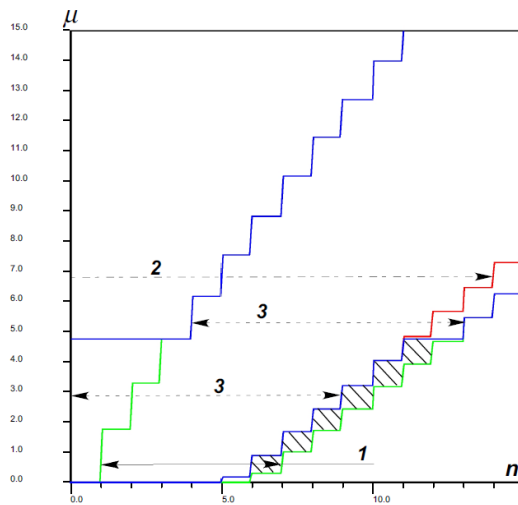


Рис. 8. Доверительные интервалы для неизвестного сигнала μ при наличии пуассоновского фона со средним $b=3$ (зелёный(1) - 90% симметричный доверительный интервал без учёта априорной информации; красный (2) - односторонний 90% интервал без учёта информации о фоне; синий (3) - 90% доверительный интервал, полученный с учётом априорной информации в оценке (estimator) параметра μ).

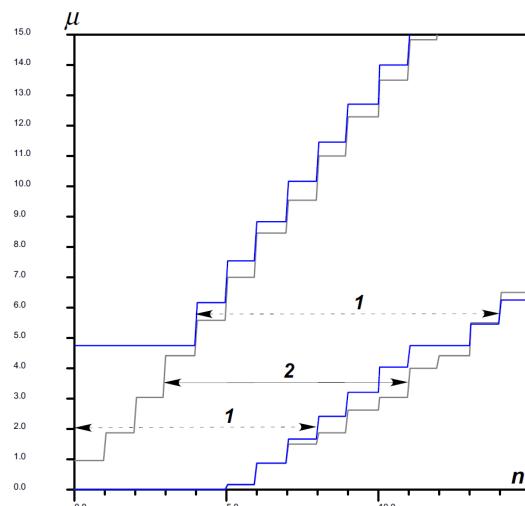


Рис. 9. Доверительные интервалы для неизвестного сигнала μ при наличии пуассоновского фона со средним $b=3$ (синий(1) - 90% доверительный интервал с учётом априорной информации (аналогично рис. 8), серый(2) - 90% доверительный интервал, построенный по методу Фельдмана и Казинса [6]).

На Рис. 9 приведены интервалы, построенные по методу Фельдмана и Казинса (серый), и интервалы, построенные с учётом априорной информации в оценке (синий). Очевидно, что доверительные интервалы с корректным учётом априорной информации (синий график)

- 1) дают корректную оценку в области $n \leq b$;
- 2) имеют участок, аналогичный участку CEF на Рис.6, дают наилучшую оценку нижней границы доверительного интервала;
- 3) по построению обеспечивают близкое к 90% вероятностное содержание доверительной области (для дискретных распределений получить 90% содержание в общем случае невозможно).

Для удобства сравнения с конструкцией [6] мы приводим Таблицу 1, аналогичную Таблицам II - IX в работе [6], в которой даются интервалы для доверительного уровня 90% и для различных значений фона и измеренного числа событий (b от 0 до 10, $n_0 = (0..20)$).

Программное обеспечение, позволяющее рассчитывать интервалы для любых значений доверительного уровня, параметров μ , b и n_0 , а также рисовать соответствующие картинки, свободно доступно с адреса [20].

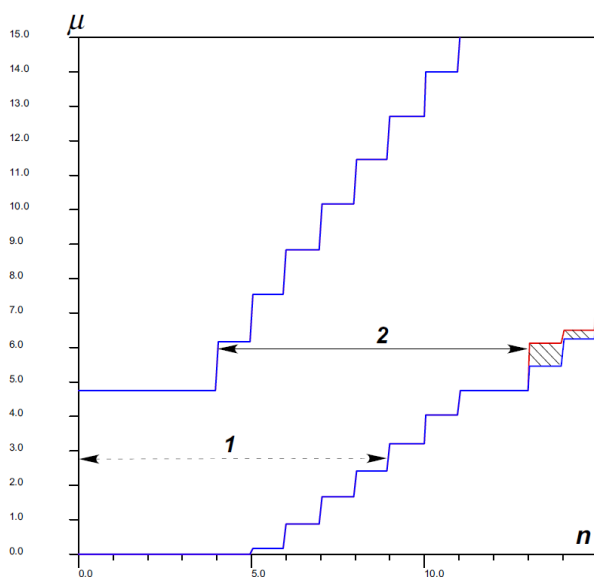


Рис. 10. Доверительные интервалы для неизвестного сигнала μ при наличии пуассоновского фона со средним $b=3$ (синий(1) - 90% доверительный интервал с учётом априорной информации, красный(2) - 90% доверительный интервал с учётом априорной информации о фоне, построенный для случая, когда важно найти ограничение сверху на исследуемый параметр μ).

Благодаря дополнительной свободе, возникающей при построении доверительных интервалов для параметров дискретных распределений (замена точного равенства (10) на \geq (11)), можно минимизировать вероятностное содержание доверительной области, приблизить его к требуемому доверительному уровню (в наших примерах - 90%). Например, на Рис. 10 область правее/ниже верхней границы удовлетворяет условию $P_{\hat{\mu}}(n_1(\mu) \geq n) \geq 0.95$. Эта верхняя граница задаёт односторонний 95% доверительный интервал. Как указывалось выше (раздел 2б), нижнюю границу можно построить формально, исходя из условия $P_{\hat{\mu}}(n_2(\mu) \leq n) \geq 0.95$ (синий график на Рис. 10). При этом заведомо выполняется условие $P_{\hat{\mu}}(n_1(\mu) \geq n \geq n_2(\mu)) \geq 0.90$.

Если верхняя граница доверительных интервалов фиксирована, то можно строить нижнюю границу $n_2(\mu)$ сразу исходя из условия $P_{\hat{\mu}}(n_1(\mu) \geq n \geq n_2(\mu)) \geq 0.90$ (красный график на Рис. 10). Вероятностное содержание доверительной области внутри красного графика будет меньше либо равно содержанию области внутри синего графика. Таким образом, отказавшись от симметричности, можно получить доверительные интервалы, более близкие к требуемому доверительному уровню по вероятностному содержанию.

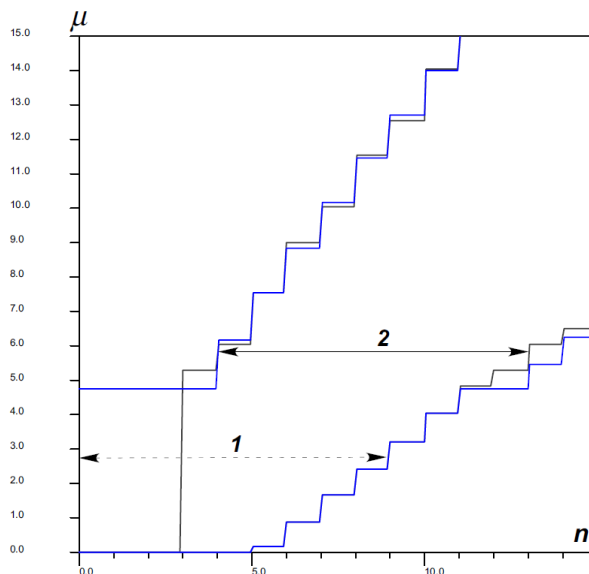


Рис. 11. Доверительные интервалы для неизвестного сигнала μ при наличии пуассоновского фона со средним $b=3$ (синий(1) - 90% доверительный интервал с учётом априорной информации, тёмно-серый(2) - 90% доверительный интервал для оценки (estimator) вида (20), построенный с помощью процедуры Стерна, Кроу и Гарднера (наиболее короткая система доверительных интервалов)).

Для полноты картины можно также сравнить (Рис. 11) доверительный интервал, построенный с учётом априорной информации (синий график), и интервал, построенный в

соответствии с процедурой Стерна, Кроу и Гарднера (темно-серый график). Построение Стерна, Кроу и Гарднера производилось для оценки вида (20). В области $n < b = 3$ серый график даёт пустой доверительный интервал, поскольку оценка (20) не принимает значений в этой области. Такие доверительные интервалы по построению имеют наименьшую длину при заданном доверительном уровне.

В результате учёта информации о фоне получено множество доверительных интервалов обладающих следующими свойствами:

- решена обозначенная в [6] проблема меньшего числа зарегистрированных событий, чем ожидаемый фон (в области, где число событий меньше фона доверительный интервал естественным образом даёт оценку сверху для измеряемого параметра μ , не зависящую от значения n , как в области $n \leq 3$ на Рис. 8);

- нижняя граница интервала имеет участок, аналогичный участку CEF на Рис.6.

- благодаря неоднозначности определения интервалов для дискретных распределений (условие (11)) возможны различные варианты построения доверительных интервалов (фиксация нижней или верхней границы, построение СКГ, их комбинации). Выбор может зависеть от конкретной ситуации, например, от необходимости получить более строгое ограничение на параметры сверху или снизу.

Таким образом, корректный учёт априорной информации о наличии фона при выборе оценки позволяет строить доверительные интервалы, отвечающие требованиям физической реалистичности и лишённые недостатков построений [6], [18] и [7]. По аналогии со случаем непрерывных распределений, рассматриваемый метод позволяет сравнивать доверительные интервалы, полученные в разных экспериментах.

8. Наилучший верхний предел (непрерывный и дискретный случаи)

а) Непрерывный случай

Рассмотрим теперь важный для экспериментальных приложений случай несимметричного доверительного интервала.

В разделе 6 обычная оценка $\hat{\theta}$ параметра θ переопределялась таким образом, что оценка учитывала априорное неравенство $\theta \geq 0$. После этого для преопределённой оценки $\tilde{\theta} = \max(\hat{\theta}, 0)$ (Ур. (16)) доверительный пояс строился обычным способом. Работа с δ -образным вкладом в распределение вероятностей для $\tilde{\theta}$ упростилась благодаря сведению задачи к построению доверительного пояса для обычной оценки $\hat{\theta}$ таким образом, чтобы итоговый пояс удовлетворял дополнительному условию (см. раздел 6). Построение было

проведено с помощью приёма так называемых горизонтальных деформаций, полученная система интервалов представлена на Рис. 6.

В разделе 6 был модифицирован стандартный симметричный доверительный пояс, соответствующий выбору $\alpha = \alpha' = (1 - \beta)/2$ в обозначениях Рис. 1. Естественным развитием данного построения будет такая же модификация для асимметричного случая $\alpha' = 0$, $\beta = 1 - \alpha$, который соответствует верхней границе для доверительного уровня β :

$$P(\theta < l_{1-\beta}(\hat{\theta})) = \beta \quad (22)$$

Такой вариант удобно использовать, если необходимо измерить положительный сигнал, в то время как точность не позволяет доказать наличие ненулевого сигнала с достаточной доверительной вероятностью. В этом случае желательно указать наиболее компактный верхний предел. Такая постановка задачи и её решение рассмотрены в работе [12].

Модифицируем доверительный интервал (22) так, чтобы он удовлетворял априорному неравенству $\theta \geq 0$.

Необходимое геометрическое построение даётся на Рис. 12, который отличается от Рис. 2. добавленными точками пересечения (точки MBDN на горизонтали LG).

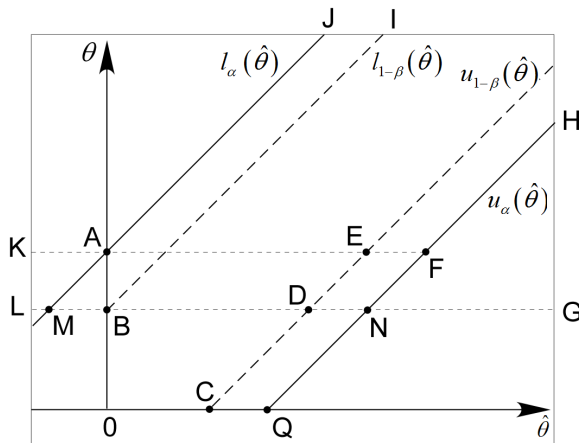


Рис. 12. Пары сплошных и пунктирных наклонных линий задают симметричные доверительные пояса для доверительных уровней $\beta = 1 - 2\alpha$ и $\tilde{\beta} = 1 - 2(1 - \beta) = 1 - 4\alpha$. Функции, соответствующие этим линиям обозначены на рисунке. Точки A и B суть пересечения кривых $\theta = l_{\alpha}(\hat{\theta})$ и $\theta = l_{1-\beta}(\hat{\theta})$ с вертикальной осью. Точки A и B задают горизонтальные линии KF и LG с другими точками пересечения.

В последующем построении будут участвовать только точки M, B, C, D, N; остальные точки указаны для явной связи с Рис. 2.

Величина θ_B есть ордината точки B (а также точек M, D, и N):

$$\theta_B = l_{1-\beta}(0) \quad (23)$$

Величины $\theta_C < \theta_B$ суть абсциссы точек C и D:

$$\theta_C = U_{1-\beta}(0), \quad \theta_D = U_{1-\beta}(\theta_B) \quad (24)$$

Немодифицированная граница (22) соответствует доверительным интервалам (с уровнем доверия β), которые начинаются на пунктирной линии BI и продолжаются вниз до бесконечности.

Для получения модификации границы (22), начнем с доверительного пояса $[u(\hat{\theta}), l(\hat{\theta})]$, обозначенного на Рис. 13 жирными линиями.

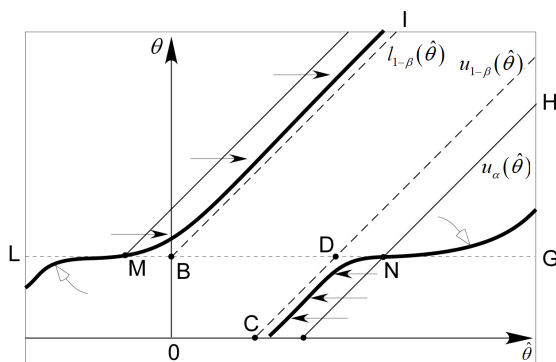


Рис. 13. Две жирных кривых задают допустимый доверительный пояс для доверительного уровня β . Жирные линии "закреплены" в точках M и N. Черные стрелки указывают допустимые горизонтальные деформации. Белыми стрелками указано распрямление соответствующих сегментов.

После этого проводим горизонтальные деформации u и l , как показано черными стрелками (подробнее приём описан в разделе 6). В результате сегмент u ниже точки N прижимается к прямой CD , одновременно часть l выше точки M прижимается к прямой BI . Деформации с другой стороны обозначены белыми стрелками. Полученный таким образом доверительный пояс показан на Рис. 14.

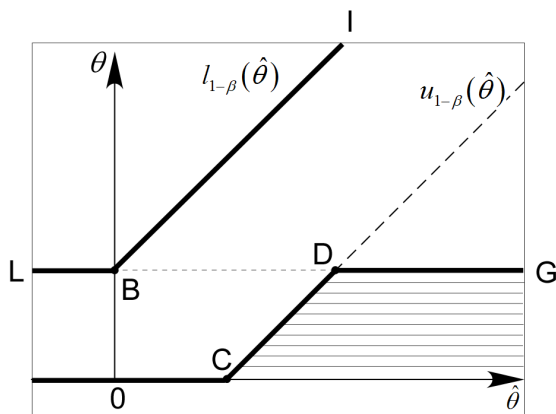


Рис. 14. Доверительный пояс, полученный из немодифицированной верхней границы для доверительного уровня β с учётом априорной информации $\theta \geq 0$. Область под ломаной CDG является выигрышем за счёт учёта априорной информации.

Аналитически построение можно описать следующим образом (мы рассматриваем доверительный уровень β):

— Для $\hat{\theta} \geq \theta_B$ доверительный интервал есть $[\theta_B, l_{1-\beta}(\hat{\theta})]$, то есть верхняя граница та же самая, что и в немодифицированном случае (Ур. (22)), но ограничен снизу величиной θ_B . Область под ломаной CDG является выигрышем за счёт учёта априорной информации.

— Для $\theta_C \leq \hat{\theta} \leq \theta_B$ доверительный интервал $[u_{1-\beta}(\hat{\theta}), l_{1-\beta}(\hat{\theta})]$, то есть в точности немодифицированный симметричный доверительный интервал для доверительного уровня $\tilde{\beta} = 1 - 2(1 - \beta) = 1 - 4\alpha$.

— Для $0 \leq \hat{\theta} \leq \theta_C$ доверительный интервал $[0, l_{1-\beta}(\hat{\theta})]$, то есть немодифицированная верхняя граница (22), ограниченная снизу физическим условием.

— Наконец, для $\hat{\theta} \leq 0$ доверительный интервал фиксирован и имеет вид $[0, \theta_B]$.

Примечательными свойствами такого доверительного пояса являются:

— Оценивание устойчиво для нефизических значений оценки параметра, то есть для $\hat{\theta} < 0$.

— Верхняя граница интервала для физических значений $\hat{\theta}$ та же, что и для немодифицированного случая (22) и является наименьшей возможной для данного доверительного уровня β .

— Нижняя граница интервала отходит от нуля в точке с наименьшей возможной (для данного доверительного уровня) абсциссой θ_C , кроме того, нижняя граница является максимально возможной для данного доверительного уровня в интервале $\theta_C \leq \hat{\theta} \leq \theta_B$.

— Для вычисления доверительного интервала не требуется сложных алгоритмов или таблиц сверх стандартных процедур построения доверительного интервала для доверительного уровня $\tilde{\beta} = 1 - 2(1 - \beta) = 1 - 4\alpha$.

— Важной особенностью такого верхнего предела оказывается отсутствие перекрывания (превышения доверительного уровня), характерного для искусственных рецептов типа [7].

б) Дискретный случай

Для параметров дискретных распределений построение верхнего предела с учётом априорной информации производится по аналогии с построением в непрерывном случае.

Оценка (estimator) выбирается в виде (20). Далее, аналогично Рис. 12 и 13, рассматриваются доверительные интервалы: 90% двусторонний без учёта априорной информации, а также верхняя и нижняя границы 90% односторонних интервалов (Рис. 15). Хотя для дискретных распределений процедура горизонтальных деформаций не определена, в результате выбора оценки (20) с учётом информации о фоне, получим доверительную область (ограничена черными графиками на Рис. 15), схожую по конфигурации с доверительным поясом на Рис. 14.

Наилучший верхний предел для неизвестного сигнала μ при наличии пуассоновского фона со средним b с учётом априорной информации о фоне в виде оценки (20) обладает следующими свойствами:

— При измерении в эксперименте любого числа событий $n \leq b$ получим одинаковый доверительный интервал. Это также решает проблему меньшего, чем фон, числа событий, обозначенную в [6].

— Вероятностное содержание области между черными графиками близко к требуемому доверительному уровню. Превышение доверительного уровня (перекрывание в понимании [6]) является недостатком некоторых других построений [19].

— Нижняя граница доверительного интервала отрывается от оси n (становится отличной от нуля) при наименьшем возможном значении n . Таким образом, не только указывается наилучший (наиболее сильный) верхний предел для параметра, но и появляется возможность детектировать сигнал (отличная от нуля нижняя граница в некотором смысле обозначает наличие сигнала) при достаточно малых значениях измеренного числа событий.

— Как и в непрерывном случае, область под нижней границей является выигрышем от использования априорной информации о фоне.

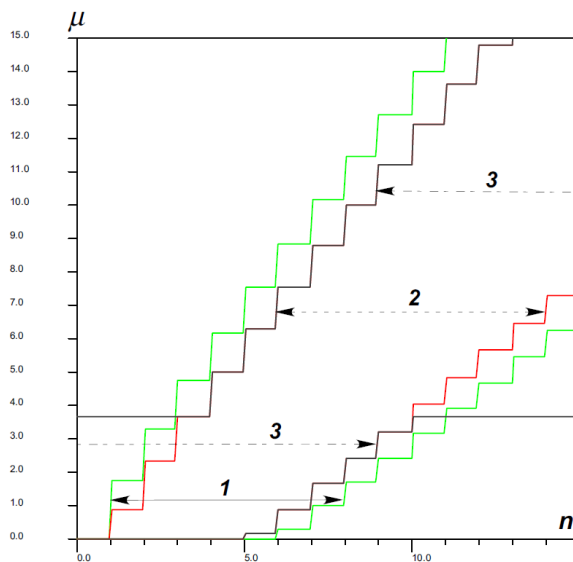


Рис. 15. Верхняя граница доверительного интервала для неизвестного сигнала μ при наличии пуассоновского фона со средним $b=3$. (зелёный(1) - 90% двусторонний доверительный интервал без учёта априорной информации, красный(2) - 90% верхний и нижние интервалы без учёта априорной информации, чёрный(3) - 90% односторонний доверительный интервал для оценки (estimator) вида (20) - наилучшая верхняя граница интервала по аналогии с [12]).

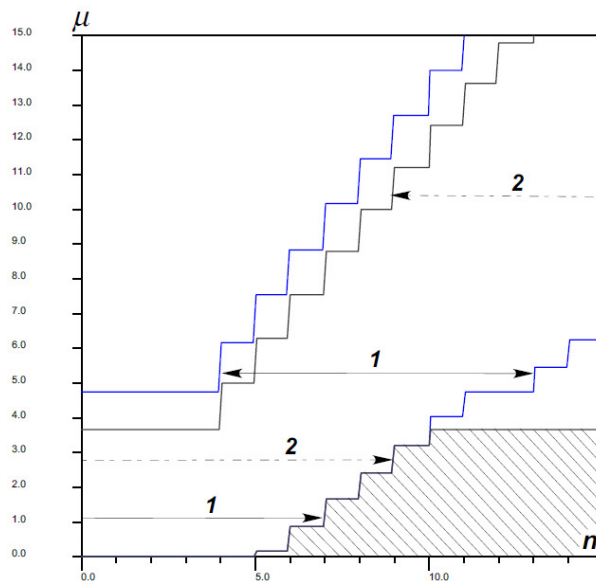


Рис. 16. Доверительные интервалы для неизвестного сигнала μ при наличии пуассоновского фона со средним $b=3$. (синий(1) - 90% двусторонний доверительный интервал с учётом априорной информации, серый(2) - 90% интервал для оценки вида (20) - наилучшая верхняя граница доверительного интервала по аналогии с [12]).

Для сравнения приведём также двусторонний (симметричный) 90% интервал с учётом априорной информации (синий график, Рис. 16) и 90% наилучший верхний предел (серый график, Рис. 16) для оценки вида (20). Видно, что наилучшая верхняя граница

доверительного интервала для параметра μ лежит ниже верхней границы двустороннего интервала. Нижние границы совпадают до точки $n = 10$. Очевидно, что двусторонний интервал даёт более универсальную оценку - для любого измеренного числа событий. Система доверительных интервалов, определяющая наилучшую верхнюю границу для параметра, по построению также существует для любого значения n . Однако, наибольший интерес такая система интервалов представляет в случае, когда n мало ($n < 10$). Именно в этой области с одной стороны верхний предел наиболее строгий, с другой стороны возможно детектирование (в описанном выше смысле: отрыв нижней границы от нуля).

Ещё раз подчеркнём, что корректное построение верхнего предела с учётом априорной информации в непрерывном и дискретном случаях по методу предела чувствительности даёт возможность осмысленно сравнивать результаты экспериментов, представленных непосредственно в виде доверительных интервалов.

9. Выводы

Таким образом, задача об учете априорной информации о параметрах при построении доверительных интервалов допускает полноценное, физически корректное решение, которое мы называем методом предела чувствительности, так как он уточняет известный рецепт предела чувствительности.

Это решение имеет прозрачное и потому убедительное обоснование, и оказывается простым и удобным для практического применения.

Перечислим важнейшие особенности рассмотренной конструкции.

Непрерывные распределения

Система доверительных интервалов, построенная на основе выбора оценки с учётом априорной информации об ограниченном параметре непрерывного распределения, обладает следующими примечательными свойствами (как для симметричного интервала, так и для наилучшего верхнего предела):

— Оценивание устойчиво для нефизических значений оценки параметра, то есть для $\hat{\theta} < 0$.

— Верхняя граница интервала для физических значений $\hat{\theta}$ та же, что и для немодифицированного случая и является наименьшей возможной для данного доверительного уровня β (для наилучшего верхнего предела). В случае симметричного интервала для физических значений $\hat{\theta}$ верхняя граница всегда лежит ниже верхней границы для построения [6].

— Нижняя граница интервала отходит от нуля в точке с наименьшей возможной (для данного доверительного уровня) абсциссой θ_c (Рис. 6 и 14), кроме того, нижняя граница является максимально возможной для данного доверительного уровня на отрезках $\theta_c \leq \hat{\theta} \leq \theta_\epsilon$ (Рис. 6), $\theta_c \leq \hat{\theta} \leq \theta_b$ (Рис. 14).

— Для вычисления доверительного интервала не требуется сложных алгоритмов или таблиц помимо стандартных процедур построения доверительного интервала для заданного доверительного уровня.

Дискретные распределения

В результате учёта информации о фоне получено множество доверительных интервалов обладающих следующими свойствами:

— Решается указанная в [6] проблема меньшего числа зарегистрированных событий, чем ожидаемый фон (в области, где число событий меньше фона, построенная система доверительных интервалов естественным образом даёт оценку сверху для измеряемого параметра, не зависящую от значения n , как в области $n \leq 3$ на Рис. 8);

— Нижняя граница интервала имеет участок, аналогичный участку CEF на Рис.6; благодаря учёту ограничения на параметр отрыв нижней границы доверительного интервала от оси n происходит раньше.

— Благодаря неоднозначности определения интервалов для дискретных распределений (условие (11)) возможны различные варианты построения доверительных интервалов (фиксация нижней или верхней границы, построение Стерна, Кроу и Гарднера, их комбинации). Выбор может зависеть от конкретной ситуации, например, от необходимости получить более строгое ограничение на параметры сверху или снизу.

Наилучший верхний предел для неизвестного сигнала μ при наличии пуассоновского фона со средним b с учётом априорной информации о фоне в виде оценки (20) обладает следующими свойствами:

— При измерении в эксперименте любого числа событий $n \leq b$ получим одинаковый доверительный интервал. Это, в частности, решает проблему меньшего, чем фон, числа событий, обозначенную в [6].

— Вероятностное содержание доверительной области близко к требуемому доверительному уровню, перекрытие мало (для дискретных распределений полностью его избежать не удаётся). Перекрытие (в понимании [6]) является недостатком некоторых других построений [19], [7].

— Нижняя граница доверительного интервала отходит от оси n (становится отличной от нуля) при наименьшем возможном значении n . Таким образом, не только

указывается наилучший (наиболее сильный) верхний предел для параметра, но и появляется возможность детектировать сигнал (отличная от нуля нижняя граница в некотором смысле обозначает наличие сигнала) при достаточно малых значениях измеренного числа событий.

— Как и в непрерывном случае, область под нижней границей является выигрышем от использования априорной информации о фоне.

Приведённое выше построение доверительных интервалов по методу предела чувствительности для непрерывного и дискретного случаев обладает рядом достоинств:

- 1) приведён явный вид модифицированных оценок и соответствующие распределения;
- 2) дана интерпретация вида распределений: указана неразличимость полученных в эксперименте оценок, лежащих вне физической области значений параметров;
- 3) по аналогии с симметричными интервалами оценивается наилучший верхний предел для параметров распределений с учётом априорной информации;
- 4) рассматриваются возможные вариации: например, построен более короткий доверительный интервал ценой отказа от симметричности (по аналогии с [5]), а также интервалы с закреплённым верхним или нижним пределами;
- 5) правильность вероятностного содержания доверительных областей, соответствующего условиям (3) и (11), достигается благодаря использованию стандартной неймановской процедуры построения доверительной области, независимо от вида оценки.

Важнейшее свойство обсуждаемого метода — возможность прямого сравнения доверительных интервалов, полученных в разных экспериментах. В отличие от прочих рецептов учёта априорной информации, метод предела чувствительности устойчив (даёт физически корректные интервалы) вблизи физической границы значений параметра.

Следует помнить, что суть и смысл обработки данных состоит в том, чтобы сама величина получаемых чисел непосредственно отражала ту информацию, которую требуется извлечь из данных посредством обработки. В этом смысле методы Фельдмана и Казинса и Кауэна и др. просто не достигают цели, можно сказать, шифруя и тем самым скрывая некоторую существенную часть информации. Метод предела чувствительности лишён этого недостатка, эффективно решая задачу построения доверительных интервалов с учётом априорной информации о параметрах.

Литература

- [1] Neyman, Outline of a theory of statistical estimation based on the classical theory of probability, Phil. Trans. R. Soc. Lond. A 1937 236, 333-380.
- [2] V.N. Aseev et al., Phys. Rev. D 84 (2011) 112003.
- [3] H.V. Klapdor-Kleingrothaus et al., Eur. Phys. J. A12 (2001) 147, E. Andreotti et al., Astropart. Phys. 34 (2011) 822, arXiv:1012.3266, R. Arnold et al., Nucl. Phys. A765 (2006) 483, hep-ex/0601021, A. Barabash et al., Phys.Atom.Nucl. 74 (2011) 312, arXiv:1002.2862.
- [4] K. Abe et al., Phys. Rev. Lett. 107 (2011) 041801, arXiv:1106.2822, P. Adamson et al., Phys. Rev. Lett. 107 (2011) 181802, arXiv:1108.0015.
- [5] T.E. Sterne, Biometrika 41 (1954) 275, E.L. Crow, Biometrika 43 (1956) 423, E.L. Crow and R.S. Gardner, Biometrika 46 (1959) 441.
- [6] G. J. Feldman and R. D. Cousins, Phys. Rev. D 57 (1998) 3873.
- [7] G. Cowan, K. Cranmer, E. Gross, O. Vitells, Power-Constrained Limits, arXiv:1105.3166.
- [8] M. Mandelkern, J. Schultz, The Statistical Analysis of Gaussian and Poisson Signals Near Physical Boundaries, J.Math.Phys. 41 (2000) 5701-5709, arXiv:hep-ex/9910041v3
- [9] F.Tkachov, Optimal confidence intervals for bounded parameters, arXiv: 0911.4271
- [10] Ch. Kraus et al., Eur.Phys.J. C 40 (2005) 447.
- [11] V.M. Lobashev, Prog.Part.Nucl.Phys. **48** (2002) 123; Nucl. Phys. A **719** (2003) 153; V.M. Lobashev et al., Nucl. Phys. B (Proc. Suppl.) **91** (2001) 280.
- [12] F.Tkachov, Optimal upper bounds for non-negative parameters, arXiv:0912.1555
- [13] А.А. Тяпкин, По поводу трактовки основных проблем теории оценок (Статистические методы в экспериментальной физике / пер. с англ. В.С. Курбатова, под ред. проф. А.А. Тяпкина; [Идье В., Драйард Д., Джеймс Ф. и др.]. - М. : Атомиздат, 1976. - 335 с.)
- [14] A. L. Read, Modified frequentist analysis of search results (the CLs method), in Proceedings of the Workshop on Confidence Limits, CERN, Switzerland, 17-18 January 2000, F. James, L. Lyons, and Y. Perrin, eds. (2000), p. 81.
- [15] The ATLAS Collaboration, Phys.Lett. B 716 (2012) 1; The CMS Collaboration, Phys.Lett. B 716 (2012) 30.
- [16] W.T. Eadie, D. Dryard, F.E. James, M. Roos and B. Sadoulet, Statistical methods in experimental physics. North-Holland, 1971.
- [17] F.V. Tkachov, Transcending the least squares. arXiv:physics/0604127
- [18] C. Giunti, Phys.Rev.D 59 (1999) 053001.
- [19] R.D. Cousins, Negatively Biased Relevant Subsets Induced by the Most-Powerful One-Sided Upper Confidence Limits for a Bounded Physical Parameter, arXiv:1109.2023.
- [20] <http://www.inr.ac.ru/~blackbox/stat/intervals/>

Таблица 1

90% доверительные интервалы с учётом априорной информации для параметра μ , при известном значении фона b в интервале от 0 до 10 и при измеренном числе событий

$n_0 = (0..20)$. (аналогично Таблицам II - IX в работе [6])

	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5
0	0.0, 3.0.	0.0, 2.5.	0.0, 3.75.	0.0, 3.25.	0.0, 4.3.	0.0, 3.8.	0.0, 4.76.	0.0, 4.26.
1	0.11, 4.75.	0.0, 4.25.	0.0, 3.75.	0.0, 3.25.	0.0, 4.3.	0.0, 3.8.	0.0, 4.76.	0.0, 4.26.
2	0.54, 6.3.	0.04, 5.8.	0.0, 5.3.	0.0, 4.8.	0.0, 4.3.	0.0, 3.8.	0.0, 4.76.	0.0, 4.26.
3	1.11, 7.76.	0.61, 7.26.	0.11, 6.76.	0.0, 6.26.	0.0, 5.76.	0.0, 5.26.	0.0, 4.76.	0.0, 4.26.
4	1.75, 9.16.	1.25, 8.66.	0.75, 8.16.	0.25, 7.66.	0.0, 7.16.	0.0, 6.66.	0.0, 6.16.	0.0, 5.66.
5	2.44, 10.52.	1.94, 10.02.	1.44, 9.52.	0.94, 9.02.	0.44, 8.52.	0.0, 8.02.	0.0, 7.52.	0.0, 7.02.
6	3.0, 11.85.	2.5, 11.35.	2.16, 10.85.	1.66, 10.35.	1.16, 9.85.	0.66, 9.35.	0.16, 8.85.	0.0, 8.35.
7	3.29, 13.15.	2.79, 12.65.	2.9, 12.15.	2.4, 11.65.	1.9, 11.15.	1.4, 10.65.	0.9, 10.15.	0.4, 9.65.
8	3.99, 14.44.	3.49, 13.94.	3.66, 13.44.	3.16, 12.94.	2.66, 12.44.	2.16, 11.94.	1.66, 11.44.	1.16, 10.94.
9	4.7, 15.71.	4.2, 15.21.	3.75, 14.71.	3.25, 14.21.	3.44, 13.71.	2.94, 13.21.	2.44, 12.71.	1.94, 12.21.
10	5.43, 16.97.	4.93, 16.47.	4.43, 15.97.	3.93, 15.47.	4.23, 14.97.	3.73, 14.47.	3.23, 13.97.	2.73, 13.47.
11	6.17, 18.21.	5.67, 17.71.	5.17, 17.21.	4.67, 16.71.	4.3, 16.21.	3.8, 15.71.	4.03, 15.21.	3.53, 14.71.
12	6.93, 19.45.	6.43, 18.95.	5.93, 18.45.	5.43, 17.95.	4.93, 17.45.	4.43, 16.95.	4.76, 16.45.	4.26, 15.95.
13	7.69, 20.67.	7.19, 20.17.	6.69, 19.67.	6.19, 19.17.	5.69, 18.67.	5.19, 18.17.	4.76, 17.67.	4.26, 17.17.
14	8.47, 21.89.	7.97, 21.39.	7.47, 20.89.	6.97, 20.39.	6.47, 19.89.	5.97, 19.39.	5.47, 18.89.	4.97, 18.39.
15	9.25, 23.1.	8.75, 22.6.	8.25, 22.1.	7.75, 21.6.	7.25, 21.1.	6.75, 20.6.	6.25, 20.1.	5.75, 19.6.
16	10.04, 24.31.	9.54, 23.81.	9.04, 23.31.	8.54, 22.81.	8.04, 22.31.	7.54, 21.81.	7.04, 21.31.	6.54, 20.81.
17	10.84, 25.5.	10.34, 25.0.	9.84, 24.5.	9.34, 24.0.	8.84, 23.5.	8.34, 23.0.	7.84, 22.5.	7.34, 22.0.
18	11.64, 26.7.	11.14, 26.2.	10.64, 25.7.	10.14, 25.2.	9.64, 24.7.	9.14, 24.2.	8.64, 23.7.	8.14, 23.2.
19	12.45, 27.88.	11.95, 27.38.	11.45, 26.88.	10.95, 26.38.	10.45, 25.88.	9.95, 25.38.	9.45, 24.88.	8.95, 24.38.
20	13.26, 29.07.	12.76, 28.57.	12.26, 28.07.	11.76, 27.57.	11.26, 27.07.	10.76, 26.57.	10.26, 26.07.	9.76, 25.57.
	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5
0	0.0, 5.16.	0.0, 4.66.	0.0, 5.52.	0.0, 5.02.	0.0, 5.85.	0.0, 5.35.	0.0, 6.15.	0.0, 5.65.
1	0.0, 5.16.	0.0, 4.66.	0.0, 5.52.	0.0, 5.02.	0.0, 5.85.	0.0, 5.35.	0.0, 6.15.	0.0, 5.65.
2	0.0, 5.16.	0.0, 4.66.	0.0, 5.52.	0.0, 5.02.	0.0, 5.85.	0.0, 5.35.	0.0, 6.15.	0.0, 5.65.
3	0.0, 5.16.	0.0, 4.66.	0.0, 5.52.	0.0, 5.02.	0.0, 5.85.	0.0, 5.35.	0.0, 6.15.	0.0, 5.65.
4	0.0, 5.16.	0.0, 4.66.	0.0, 5.52.	0.0, 5.02.	0.0, 5.85.	0.0, 5.35.	0.0, 6.15.	0.0, 5.65.
5	0.0, 6.52.	0.0, 6.02.	0.0, 5.52.	0.0, 5.02.	0.0, 5.85.	0.0, 5.35.	0.0, 6.15.	0.0, 5.65.
6	0.0, 7.85.	0.0, 7.35.	0.0, 6.85.	0.0, 6.35.	0.0, 5.85.	0.0, 5.35.	0.0, 6.15.	0.0, 5.65.
7	0.0, 9.15.	0.0, 8.65.	0.0, 8.15.	0.0, 7.65.	0.0, 7.15.	0.0, 6.65.	0.0, 6.15.	0.0, 5.65.
8	0.66, 10.44.	0.16, 9.94.	0.0, 9.44.	0.0, 8.94.	0.0, 8.44.	0.0, 7.94.	0.0, 7.44.	0.0, 6.94.
9	1.44, 11.71.	0.94, 11.21.	0.44, 10.71.	0.0, 10.21.	0.0, 9.71.	0.0, 9.21.	0.0, 8.71.	0.0, 8.21.
10	2.23, 12.97.	1.73, 12.47.	1.23, 11.97.	0.73, 11.47.	0.23, 10.97.	0.0, 10.47.	0.0, 9.97.	0.0, 9.47.
11	3.03, 14.21.	2.53, 13.71.	2.03, 13.21.	1.53, 12.71.	1.03, 12.21.	0.53, 11.71.	0.03, 11.21.	0.0, 10.71.
12	3.83, 15.45.	3.33, 14.95.	2.83, 14.45.	2.33, 13.95.	1.83, 13.45.	1.33, 12.95.	0.83, 12.45.	0.33, 11.95.
13	4.65, 16.67.	4.15, 16.17.	3.65, 15.67.	3.15, 15.17.	2.65, 14.67.	2.15, 14.17.	1.65, 13.67.	1.15, 13.17.
14	5.16, 17.89.	4.66, 17.39.	4.47, 16.89.	3.97, 16.39.	3.47, 15.89.	2.97, 15.39.	2.47, 14.89.	1.97, 14.39.
15	5.25, 19.1.	4.75, 18.6.	5.3, 18.1.	4.8, 17.6.	4.3, 17.1.	3.8, 16.6.	3.3, 16.1.	2.8, 15.6.
16	6.04, 20.31.	5.54, 19.81.	5.52, 19.31.	5.02, 18.81.	5.14, 18.31.	4.64, 17.81.	4.14, 17.31.	3.64, 16.81.
17	6.84, 21.5.	6.34, 21.0.	5.84, 20.5.	5.34, 20.0.	5.85, 19.5.	5.35, 19.0.	4.98, 18.5.	4.48, 18.0.
18	7.64, 22.7.	7.14, 22.2.	6.64, 21.7.	6.14, 21.2.	5.85, 20.7.	5.35, 20.2.	5.83, 19.7.	5.33, 19.2.
19	8.45, 23.88.	7.95, 23.38.	7.45, 22.88.	6.95, 22.38.	6.45, 21.88.	5.95, 21.38.	6.15, 20.88.	5.65, 20.38.
20	9.26, 25.07.	8.76, 24.57.	8.26, 24.07.	7.76, 23.57.	7.26, 23.07.	6.76, 22.57.	6.26, 22.07.	5.76, 21.57.

	8.0	8.5	9.0	9.5	10.0
0	0.0, 6.44.	0.0, 5.94.	0.0, 6.71.	0.0, 6.21.	0.0,6.97.
1	0.0, 6.44.	0.0, 5.94.	0.0, 6.71.	0.0, 6.21.	0.0,6.97.
2	0.0, 6.44.	0.0, 5.94.	0.0, 6.71.	0.0, 6.21.	0.0,6.97.
3	0.0, 6.44.	0.0, 5.94.	0.0, 6.71.	0.0, 6.21.	0.0,6.97.
4	0.0, 6.44.	0.0, 5.94.	0.0, 6.71.	0.0, 6.21.	0.0,6.97.
5	0.0, 6.44.	0.0, 5.94.	0.0, 6.71.	0.0, 6.21.	0.0,6.97.
6	0.0, 6.44.	0.0, 5.94.	0.0, 6.71.	0.0, 6.21.	0.0,6.97.
7	0.0, 6.44.	0.0, 5.94.	0.0, 6.71.	0.0, 6.21.	0.0,6.97.
8	0.0, 6.44.	0.0, 5.94.	0.0, 6.71.	0.0, 6.21.	0.0,6.97.
9	0.0, 7.71.	0.0, 7.21.	0.0, 6.71.	0.0, 6.21.	0.0,6.97.
10	0.0, 8.97.	0.0, 8.47.	0.0, 7.97.	0.0, 7.47.	0.0,6.97.
11	0.0, 10.21.	0.0, 9.71.	0.0, 9.21.	0.0, 8.71.	0.0,8.21.
12	0.0, 11.45.	0.0,10.95.	0.0,10.45.	0.0, 9.95.	0.0,9.45.
13	0.65,12.67.	0.15,12.17.	0.0,11.67.	0.0, 11.17.	0.0,10.67.
14	1.47,13.89.	0.97,13.39.	0.47,12.89.	0.0, 12.39.	0.0,11.89.
15	2.3, 15.1.	1.8, 14.6.	1.3, 14.1.	0.8, 13.6.	0.3,13.1.
16	3.14,16.31.	2.64,15.81.	2.14,15.31.	1.64,14.81.	1.14,14.31.
17	3.98, 17.5.	3.48,17.0.	2.98,16.5.	2.48, 16.0.	1.98,15.5.
18	4.83, 18.7.	4.33,18.2.	3.83,17.7.	3.33, 17.2.	2.83,16.7.
19	5.68,19.88.	5.18,19.38.	4.68,18.88.	4.18,18.38.	3.68,17.88.
20	6.44,21.07.	5.94,20.57.	5.53,20.07.	5.03,19.57.	4.53,19.07.