

Институт ядерной физики им. Г.И.Будкера

А.Д.Букин

**Сравнение метода
построения доверительного интервала
по функции правдоподобия
с классическим методом Неймана**

Budker INP 2001–12

Новосибирск
2001

Сравнение метода построения доверительного интервала по функции правдоподобия с классическим методом Неймана

А.Д.Букин

Институт ядерной физики им. Г.И.Будкера,
630090, Новосибирск, РФ

Аннотация

В работе для нескольких примеров из экспериментальной физики элементарных частиц проведен расчет методом Монте-Карло вероятности покрытия истинных значений параметров теоретической модели доверительными интервалами, оцененными по логарифмической функции правдоподобия. При малой статистике, обычно, появляется заметное отклонение этой величины от ожидаемого уровня доверия.

Для нескольких простых случаев вычислены доверительные интервалы по классическому методу Неймана, которые могут быть использованы в соответствующих случаях вместо приближенных оценок по функции правдоподобия.

Проведено сравнение нескольких систем доверительных интервалов для параметра Пуассонова распределения. В результате представляется необоснованным переход в 1998 г. в рекомендациях Particle Data Group от верхнего предела $p < 2.30$ к $p < 2.44$ на 90%-ом уровне доверия при нулевом числе зарегистрированных событий.

Comparison of likelihood function method and classical Neyman's method for the construction of confidence interval

A.D.Bukin

Budker Institute of Nuclear Physics,
630090, Novosibirsk, Russia

Abstract

In the paper for several examples of elementary particle physics experiments the Monte Carlo evaluation of the probability of coverage of the true values of theoretical model parameters by loglikelihood function confidence intervals is carried out. Usually for low statistics there is noticeable deviation of this probability from expected confidence level.

For several simple cases there are calculated the confidence intervals according to the classical Neyman's method, which can be used in appropriate cases instead of the likelihood function estimates.

A comparison of several systems of the confidence intervals for Poisson distribution parameter is performed. As a result the change of Particle Data Group in 1998 of the upper limit $p < 2.30$ to $p < 2.44$ at 90% confidence level for zero number of detected events seems to be not enough founded.

Оглавление

1	Введение	5
2	Общие принципы построения доверительных интервалов	5
3	Непрерывные распределения	7
3.1	Определение среднего для Гауссова распределения	7
3.2	Измерение времени жизни	8
3.3	Измерение параметров Гауссова распределения	13
4	Дискретные распределения	18
4.1	Распределение Пуассона	18
4.2	Верхний предел при малом числе событий без вычитания фона	22
4.3	Подгонка резонансной кривой	27
5	Заключение	30

1. Введение

В экспериментальной физике элементарных частиц в настоящее время широко распространен метод построения доверительных интервалов для параметров теоретической модели с помощью функции правдоподобия [1–4]. С другой стороны, известно, что метод максимального правдоподобия дает наилучшие результаты при большой статистике, с уменьшением же статистики результаты становятся все более приближенными.

Классический метод Неймана [5] построения доверительных интервалов позволяет в большинстве случаев точно решить задачу оценки неизвестного параметра в виде интервала, однако, этот метод является гораздо более трудоемким.

В данной работе на примере нескольких распространенных задач в физике элементарных частиц исследуются границы применимости метода построения доверительных интервалов с помощью функции правдоподобия.

2. Общие принципы построения доверительных интервалов

Обычно результат эксперимента по измерению параметра a теоретической модели записывают в виде $a = p_{-\Delta p^{(-)}}^{+\Delta p^{(+)}}$. Ошибки измерения параметра $\Delta p^{(\pm)}$ не могут характеризовать среднеквадратичный разброс оценок параметра при многократном повторении этого эксперимента в одинаковых условиях. Во-первых, как правило, ошибки несимметричны ($\Delta p^{(+)} \neq \Delta p^{(-)}$), во-вторых, в общем случае дисперсия распределения вероятностей оценок параметра зависит от этого неизвестного параметра. На самом деле в этой записи содержится двойная оценка параметра: p является асимптотически несмещенной и эффективной оценкой значения параметра по методу максимального правдоподобия, а интервал $(p - \Delta p^{(-)}, p + \Delta p^{(+)})$ оценивает значение по вероятности [1–4] с некоторым уровнем доверия β (обычно $\beta = 0.683$). Это нельзя понимать так, что истинное значение параметра с вероятностью β попадает в интервал, полученный в данном эксперименте — у истинного значения нет

распределения вероятностей. Такая оценка означает, что многократно повторяемый один и тот же эксперимент будет получать *разные оценки* параметра в виде доверительного интервала, и *какое бы ни было истинное значение* параметра, оно с вероятностью β будет покрываться доверительными интервалами.

Алгоритм оценки параметра методом максимального правдоподобия следующий. Записывается логарифмическая функция правдоподобия $L(a)$, которая является логарифмом (с обратным знаком) плотности вероятности получить данный набор экспериментальных данных при некотором значении параметра a теоретической модели (или набора параметров). При вычислении доверительного интервала по какому-либо параметру модели, надо при каждом значении интересующего параметра a минимизировать функцию L по всем остальным параметрам теоретической модели. Этой процедурой функция L превращается в функцию одного аргумента. Теперь значение параметра $a = p$, при котором достигается минимум

$$L_0 = \min_a L(a) = L(p), \quad (1)$$

является оптимальной оценкой параметра по методу максимального правдоподобия, а интервал (p_1, p_2) , где

$$L(p_1) = L(p_2) = L_0 + \Delta L, \quad p_1 < p < p_2 \quad (2)$$

называется доверительным интервалом на уровне достоверности β . Величины β и ΔL связаны соотношением

$$\frac{1}{\sqrt{2\pi}} \int_{-\sqrt{2\Delta L}}^{\sqrt{2\Delta L}} e^{-\frac{x^2}{2}} dx = \beta. \quad (3)$$

Например, $\Delta L = 0.5$ соответствует $\beta \approx 0.683$, $\Delta L \approx 1.353$ соответствует $\beta = 0.9$, $\Delta L = 2$ соответствует $\beta \approx 0.9545$, $\Delta L \approx 3.318$ соответствует $\beta = 0.99$.

Традиционная запись результата со статистическими ошибками получается, если обозначить

$$\Delta p^{(+)} = p_2 - p, \quad \Delta p^{(-)} = p - p_1.$$

Такой доверительный интервал называется "центральным" по той причине, что при многократном повторении эксперимента в $(1 - \beta)/2$ доле случаев истинное значение параметра $a < p_1$, а в $(1 - \beta)/2$ доле случаев истинное значение параметра $a > p_2$.

Иногда оценка параметра статистически согласуется с физическим пределом для измеряемого параметра модели. Тогда вместо ”центрального” доверительного интервала приводят верхний предел — односторонний доверительный интервал. Экспериментальный результат в этом случае выглядит так

$$a < p_2 \text{ на уровне достоверности } \beta, \quad (4)$$

где процедура определения границы p_2 такая же, как в случае ”центрального” доверительного интервала, но соответствие между ΔL и β определяется соотношением

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sqrt{2\Delta L}} e^{-\frac{x^2}{2}} dx = \beta, \quad (5)$$

что дает для $\beta = 0.9$ $\Delta L \approx 0.822$, для $\beta = 0.99$ — $\Delta L \approx 2.707$.

Как уже было сказано, метод максимального правдоподобия дает приближенную оценку доверительного интервала. Классический метод Неймана описан ниже на примере оценки доверительного интервала при измерении времени жизни (раздел 3.2).

3. Непрерывные распределения

3.1. Определение среднего для Гауссова распределения

Задача определения среднего значения величины, экспериментальные оценки которой подчиняются Гауссову распределению с известным среднеквадратичным отклонением, является редким примером, когда метод максимального правдоподобия дает абсолютно точный результат. В общем виде задача формулируется следующим образом. В n экспериментах измерена одна и та же величина x . В каждом эксперименте своя точность измерения — отклонение от среднего подчиняется распределению Гаусса и характеризуется среднеквадратичным отклонением σ_i . Найти оптимальную оценку параметра и 68%-ый доверительный интервал.

Функция правдоподобия равна

$$L(x) = \sum_{i=1}^n \left[\frac{(x_i - x)^2}{2\sigma_i^2} - \ln(\sqrt{2\pi}\sigma_i) \right]. \quad (6)$$

Введем обозначение:

$$\sigma = \left[\sum_{i=1}^n \frac{1}{\sigma_i^2} \right]^{-\frac{1}{2}} \quad (7)$$

Легко убедиться, что минимум функции правдоподобия достигается при $x = a$

$$a = \sigma^2 \cdot \left[\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right], \quad (8)$$

а 68%-ый доверительный интервал равен $(a - \sigma, a + \sigma)$. Легко убедиться (например, методом Монте-Карло), что при многократном повторении эксперимента истинное значение среднего покрывается доверительным интервалом ровно в 68% случаев.

3.2. Измерение времени жизни

Результаты данного раздела применимы к обработке данных любого эксперимента, где измеряемая величина имеет экспоненциальное распределение вероятностей:

$$\frac{dW}{dt} = \frac{1}{\tau} \exp\left(-\frac{t}{\tau}\right) \quad (9)$$

Это может быть, например, распад нестабильной частицы или пробег частицы при постоянном коэффициенте поглощения. Для удобства далее измеряемую величину t будем называть временем жизни, а параметр теоретической модели τ — средним временем жизни.

Логарифмическая функция правдоподобия при n независимых измерениях равна:

$$L = n \cdot \ln \tau + \frac{1}{\tau} \sum_{i=1}^n t_i \quad (10)$$

Минимум этой функции L_0 достигается при

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^n t_i; \quad \min L = L_0 = n + n \cdot \ln \hat{\tau}. \quad (11)$$

Эта оценка, очевидно, является несмещённой при любом числе измерений n . Легко заметить, что статистические свойства оценки параметра и доверительных интервалов здесь не зависят от величины τ , а зависят только от числа измерений n . Границы доверительного интервала могут

Таблица 1: Границы $\Delta\tau = \tau - \hat{\tau}$ доверительных интервалов для уровня доверия 68% ($\Delta L = L - L_0 = 1/2$) и 95% ($\Delta L = 2$), определённых по методу максимального правдоподобия.

n	$\Delta L = \frac{1}{2}$			$\Delta L = 2$		
	$\frac{\Delta\tau(-)}{\hat{\tau}}$	$\frac{\Delta\tau(+)}{\hat{\tau}}$	$\sqrt{\frac{2\Delta L}{n}}$	$\frac{\Delta\tau(-)}{\hat{\tau}}$	$\frac{\Delta\tau(+)}{\hat{\tau}}$	$\sqrt{\frac{2\Delta L}{n}}$
1	0.5759	2.3144	1.0000	0.7780	18.0588	2.0000
2	0.4689	1.2283	0.7071	0.6822	5.3054	1.4142
3	0.4095	0.8940	0.5774	0.6206	3.1642	1.1547
4	0.3697	0.7253	0.5000	0.5759	2.3144	1.0000
5	0.3405	0.6212	0.4472	0.5410	1.8579	0.8944
6	0.3177	0.5496	0.4082	0.5127	1.5714	0.8165
7	0.2992	0.4968	0.3780	0.4891	1.3736	0.7559
8	0.2839	0.4560	0.3536	0.4689	1.2283	0.7071
9	0.2708	0.4232	0.3333	0.4513	1.1164	0.6667
10	0.2595	0.3963	0.3162	0.4357	1.0274	0.6325
20	0.1939	0.2614	0.2236	0.3405	0.6212	0.4472
50	0.1290	0.1558	0.1414	0.2367	0.3456	0.2828

быть определены только численными методами. В табл. 1 приведены границы доверительных интервалов для нескольких значений n и для уровня доверия 68% ($\Delta L = 1/2$) и 95% ($\Delta L = 2$). В пределе больших n доверительный интервал становится симметричным с полуразмером:

$$\Delta\tau = \hat{\tau} \cdot \sqrt{\frac{2\Delta L}{n}} \quad (12)$$

Из таблицы видно, что даже при $n = 50$ доверительные интервалы ещё не становятся симметричными. Посмотрим, насколько соответствует реальная вероятность "покрытия" истинного значения доверительным интервалом ожидаемой вероятности (рис. 1). Несмотря на небольшую величину отклонения реальной вероятности покрытия истинного значения от планируемой величины β , все же эта величина явно меньше ожидаемой при числе измерений $n < 10$ и стремится к β при $n \rightarrow \infty$. Наиболее заметно это отклонение для одностороннего доверительного интервала (верхнего предела). Здесь отклонение от β заметно даже при $n = 50$. Рассмотрим на примере такого типа экспериментов классический метод Неймана [5] построения доверительных интервалов, для которых вероятность покрытия истинного значения параметра в точности равна β .

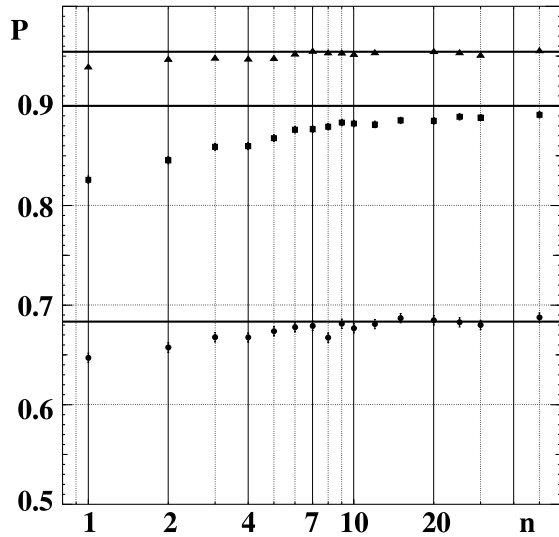


Рис. 1: Вероятность попадания истинного значения времени жизни внутрь доверительного интервала. \bullet — $\Delta L = 1/2$, \blacktriangle — $\Delta L = 2$, \blacksquare — $\Delta L = 0.8221$ (полу-бесконечный интервал, 90%-ый верхний предел). Горизонтальные сплошные линии показывают “правильное” значение вероятности (0.6827, 0.9545 и 0.9).

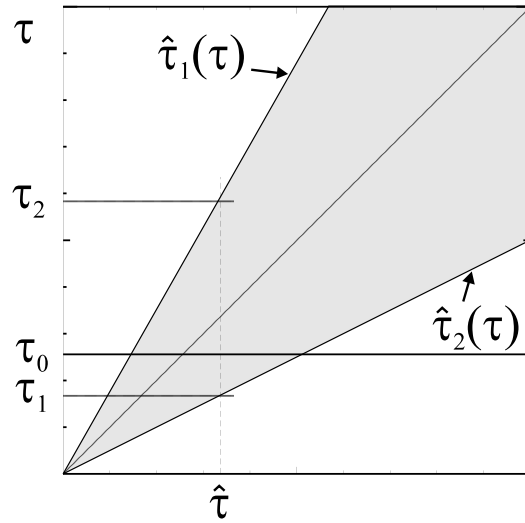


Рис. 2: Пример построения доверительной зоны и доверительных интервалов для эксперимента по измерению времени жизни с $n = 5$, $\beta = 0.6827$.

Описываемый метод справедлив для любых оценок параметров — не только для тех, что получены методом максимального правдоподобия (даже для смещённых оценок — в некоторых случаях может оказаться, что доверительный интервал не включает полученную оценку параметра).

Возьмём оценку времени жизни $\hat{\tau}$ из (11). Распределение вероятностей по значениям $\hat{\tau}$ при некотором истинном значении времени жизни τ равно

$$f_n(\hat{\tau}) = \frac{dW}{d\hat{\tau}} = \frac{1}{(n-1)!} \cdot \left(\frac{n\hat{\tau}}{\tau}\right)^{n-1} \cdot \frac{n}{\tau} \cdot \exp\left(-\frac{n\hat{\tau}}{\tau}\right) \quad (13)$$

Как уже указывалось, эта оценка параметра является несмещённой, т.е. среднее значение $\hat{\tau}$ равно τ . Наиболее вероятным значением $\hat{\tau}$ является

$$\hat{\tau}_{\text{н.в.}} = \frac{n-1}{n}\tau.$$

Используя распределение вероятностей (13), можно получить доверительный интервал для $\hat{\tau}$ при любом истинном значении τ с уровнем доверия $\beta < 1$. Центральный доверительный интервал можно определить двумя функциями $\hat{\tau}_1(\tau)$ и $\hat{\tau}_2(\tau)$ таким образом:

$$\int_0^{\hat{\tau}_1} f_n(\hat{\tau}) d\hat{\tau} = \frac{1-\beta}{2}; \quad \int_{\hat{\tau}_2}^{\infty} f_n(\hat{\tau}) d\hat{\tau} = \frac{1-\beta}{2}. \quad (14)$$

Для данной функции распределения вероятностей интеграл вычисляется в виде

$$\int_a^{\infty} f_n(\hat{\tau}) d\hat{\tau} = \left\{ 1 + \frac{na}{\tau} + \frac{1}{2} \cdot \left(\frac{na}{\tau}\right)^2 + \dots + \frac{1}{k!} \cdot \left(\frac{na}{\tau}\right)^k + \dots + \frac{1}{(n-1)!} \cdot \left(\frac{na}{\tau}\right)^{n-1} \right\} \cdot \exp\left(-\frac{na}{\tau}\right). \quad (15)$$

К сожалению, функции $\hat{\tau}_1(\tau)$ и $\hat{\tau}_2(\tau)$, найденные из уравнения (14), сложным образом зависят от n . Тем не менее, если решить уравнение для каждого τ и нанести на график, где по одной оси откладывается $\hat{\tau}$, а по другой — τ , то получится доверительная зона. На рис.2 приведён пример такой доверительной зоны для $n = 5$, $\beta = 0.6827$. Теперь очевидно, что если мы для каждой оценки $\hat{\tau}$, полученной в соответствии с формулой (11), проведём вертикальную линию до пересечения с доверительной зоной (как показано на рисунке), то получим доверительный интервал (τ_1, τ_2) , который ровно в 68.27% случаев покрывает истинное значение

Таблица 2: Границы $\Delta\tau^{(\pm)} = |\tau_{2,1} - \hat{\tau}|$ доверительных интервалов для уровня доверия 68% и 95%, определённых по классическому методу Неймана (в скобках указаны границы доверительных интервалов из таблицы 1, приближённо оцененные по функции правдоподобия).

n	$\beta = 0.6827$		$\beta = 0.9545$	
	$\frac{\Delta\tau^{(-)}}{\hat{\tau}}$	$\frac{\Delta\tau^{(+)}}{\hat{\tau}}$	$\frac{\Delta\tau^{(-)}}{\hat{\tau}}$	$\frac{\Delta\tau^{(+)}}{\hat{\tau}}$
1	0.457 (0.577)	4.789 (2.314)	0.736 (0.778)	42.45 (18.06)
2	0.394 (0.469)	1.824 (1.228)	0.648 (0.682)	7.690 (5.305)
3	0.353 (0.410)	1.194 (0.894)	0.592 (0.621)	4.031 (3.164)
4	0.324 (0.370)	0.918 (0.725)	0.551 (0.576)	2.781 (2.314)
5	0.302 (0.341)	0.760 (0.621)	0.519 (0.541)	2.159 (1.858)
6	0.284 (0.318)	0.657 (0.550)	0.492 (0.513)	1.786 (1.571)
7	0.270 (0.299)	0.584 (0.497)	0.470 (0.489)	1.538 (1.374)
8	0.257 (0.284)	0.529 (0.456)	0.452 (0.469)	1.359 (1.228)
9	0.247 (0.271)	0.486 (0.423)	0.435 (0.451)	1.225 (1.116)
10	0.237 (0.260)	0.451 (0.396)	0.421 (0.436)	1.119 (1.027)
20	0.182 (0.194)	0.285 (0.261)	0.331 (0.341)	0.654 (0.621)
50	0.124 (0.129)	0.164 (0.156)	0.232 (0.237)	0.356 (0.346)

τ_0 (какое бы ни было истинное значение!). Стоит ещё раз подчеркнуть специфику этого подхода: построение доверительной зоны идёт “по горизонтали”, а считывается доверительный интервал для каждой полученной в эксперименте оценки параметра — “по вертикали”. Нетрудно заметить, что несмотря на необходимость решения уравнения численным способом, решение для доверительного интервала в данном случае может быть представлено в удобном, не зависящем от $\hat{\tau}$, виде $\Delta\tau^{(\pm)}/\hat{\tau}$ (как в таблице 1). В табл. 2 приведены доверительные интервалы, полученные в классическом подходе. Совпадения доверительных интервалов, полученных приближённо с помощью функции правдоподобия, с теми, которые получены более точным классическим методом Неймана, нет даже при статистике $n = 50$ измерений. Однако, легко заметить, что длина интервала при больших n совпадает лучше, чем его положение. При $n = 50$ длины этих интервалов совпадают с точностью 1% (при смещении около 2% от длины интервала). Из сравнения результатов точного метода оценки доверительного интервала и приближённого по функции правдоподобия можно сделать вывод, что определение доверительного интервала по функции правдоподобия при анализе одного измерения времени

жизни (или любого другого параметра, который связан в эксперименте с экспоненциальным распределением вероятности) приводит к значительной ошибке, но с повышением статистики измерений использование приближённого метода всё более обоснованно.

При малом числе измерений n статистическая точность результата низкая, и иногда имеет смысл говорить о верхнем пределе. Под верхним пределом понимается граница одностороннего интервала, вероятность покрытия которым истинного значения измеряемого параметра называется уровнем доверия. Обычно в физике элементарных частиц используется верхний предел с 90%-ным уровнем доверия. Степень неточности приближённой оценки верхнего предела времени жизни по функции правдоподобия видна из рис.1 (результаты нанесены символами ■). Видно, что при малом числе измерений, когда этот результат и представляет интерес, ошибка в уровне доверия значительная, составляет около 10%.

Классический подход применим и для точного установления верхнего предела. Практическое отличие от нахождения двустороннего доверительного интервала состоит в следующем. При расчёте доверительной зоны надо при каждом предполагаемом значении времени жизни τ вычислять границу интервала $\hat{\tau}_1(\tau)$, для которой вероятность $P(\hat{\tau} \geq \hat{\tau}_1)$ равна заданному уровню доверия (например, 0.9). Проведя соответствующие вычисления, получим в случае однократного измерения времени жизни $\hat{\tau}$ 90%-ый верхний предел, равный $9.49 \cdot \hat{\tau}$, а для среднего по двум измерениям — $3.76 \cdot \hat{\tau}$. Метод максимального правдоподобия даёт оценку доверительного интервала $8.49 \cdot \hat{\tau}$ и $2.76 \cdot \hat{\tau}$ для одного и двух измерений, соответственно.

3.3. Измерение параметров Гауссова распределения

Рассмотрим задачу определения параметров a и σ Гауссова распределения

$$\frac{dW}{dx} = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x-a)^2}{2\sigma^2} \right] \quad (16)$$

по выборке n измерений случайной величины x_i , $i = 1, \dots, n$.

Логарифмическая функция правдоподобия зависит от двух параметров

$$L = \sum_{i=1}^n \left[\frac{(x_i - a)^2}{2\sigma^2} + \ln \left(\sqrt{2\pi}\sigma \right) \right] \quad (17)$$

и достигает минимума при

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n x_i; \quad \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{a})^2}. \quad (18)$$

Границы доверительных интервалов для ΔL , соответствующего уровню доверия β , получаются из уравнений

$$\left(\frac{a - \hat{a}}{\hat{\sigma}}\right)^2 = e^{\frac{2\Delta L}{n}} - 1; \quad 2 \ln \left(\frac{\sigma}{\hat{\sigma}}\right) + \left(\frac{\hat{\sigma}}{\sigma}\right)^2 - 1 = \frac{2\Delta L}{n}. \quad (19)$$

Первое из этих уравнений определяет симметричный доверительный интервал

$$\Delta a^{(\pm)} = \hat{\sigma} \cdot \sqrt{e^{\frac{2\Delta L}{n}} - 1}. \quad (20)$$

Второе уравнение определяет несимметричный доверительный интервал $\Delta \sigma^{(\pm)}$ и может быть решено только численно для конкретных ΔL и n .

Обобщение классического метода Неймана построения доверительных интервалов на случай зависимости теоретической модели от нескольких параметров легче сделать в терминах "эллипсов ошибок", когда доверительные интервалы указываются не для каждого параметра в отдельности, а в виде области в многомерном пространстве (хотя технические сложности могут сделать это практически невозможным). Все же попробуем определить доверительные интервалы отдельно для a и σ .

Известно, что переменная

$$t = \sqrt{n-1} \cdot \frac{\hat{a} - a}{\hat{\sigma}} \quad (21)$$

имеет стандартное распределение Стьюдента (см., например, [1]), не зависящее от значений a и σ . Плотность вероятности распределения Стьюдента с $(n-1)$ степенью свободы равна

$$\frac{dW}{dt} = \frac{A}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}}, \quad -\infty < t < +\infty. \quad (22)$$

Здесь A — нормировочная константа. Пользуясь этим распределением, для каждого n и уровня доверия β можно посчитать границы симметричного доверительного интервала для переменной t : $-\Delta t < t < \Delta t$. Отсюда

можно определить симметричный доверительный интервал для параметра a :

$$\Delta a = |a - \hat{a}| < \hat{\sigma} \cdot \frac{\Delta t}{\sqrt{n-1}} = \frac{\Delta t}{\sqrt{n(n-1)}} \sqrt{\sum_{i=1}^n (x_i - \hat{a})^2}, \quad (23)$$

который независимо от истинных значений a и σ в β доле случаев будет включать истинное значение a .

Распределение вероятностей для величины $\mu = n \cdot \hat{\sigma}^2 / \sigma^2$ является стандартным распределением χ^2 с $(n-1)$ степенью свободы. Для него можно найти границы доверительного интервала $\mu_1 < \mu < \mu_2$ с уровнем доверия β . Тогда границы доверительного интервала для параметра σ получаются

$$\sigma_1 = \sqrt{\frac{n}{\mu_2}} \cdot \hat{\sigma} < \sigma < \sqrt{\frac{n}{\mu_1}} \cdot \hat{\sigma} = \sigma_2. \quad (24)$$

В таблице 3 приведены границы доверительных интервалов для a и σ при нескольких значениях n . Вместо оценки максимального правдоподобия $\hat{\sigma}$ на практике чаще используется так называемая “несмещенная” оценка среднеквадратичного отклонения

$$\tilde{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{a})^2} \quad (25)$$

Как видно из таблицы 4, обе эти оценки являются смещенными, несмещенной оценкой является только оценка дисперсии $\tilde{\sigma}^2$. Так как границы доверительных интервалов не зависят от вспомогательных обозначений, то при вычислении границ вместо $\hat{\sigma}$ можно подставлять ее выражение через $\tilde{\sigma}$:

$$\hat{\sigma} = \sqrt{\frac{n-1}{n}} \cdot \tilde{\sigma}.$$

Отличие приближенной оценки границ доверительных интервалов от точных в таблице 3 заметны, но какая при этом допускается ошибка в вероятности покрытия истинного значения из этой таблицы не видно. Для определения этой вероятности были проведены расчеты методом Монте-Карло, и результаты представлены на рис.3 для среднего и среднеквадратичного отклонения отдельно.

Таблица 3: Границы доверительных интервалов для параметров Гауссова распределения, определяемых по n измерениям переменной x . В скобках указаны оценки границ доверительных интервалов по функции правдоподобия по формуле (19).

n	$\beta = 0.6827 (\Delta L = 0.5)$		
	$\frac{\Delta a}{\hat{\sigma}}$	$\frac{\sigma_1}{\hat{\sigma}}$	$\frac{\sigma_2}{\hat{\sigma}}$
2	1.837 (0.805)	1.003 (0.651)	7.065 (1.821)
3	0.934 (0.629)	0.903 (0.698)	2.947 (1.604)
4	0.691 (0.533)	0.878 (0.729)	2.190 (1.493)
5	0.571 (0.471)	0.870 (0.751)	1.879 (1.424)
6	0.497 (0.426)	0.868 (0.768)	1.708 (1.376)
7	0.445 (0.392)	0.869 (0.782)	1.600 (1.341)
8	0.407 (0.365)	0.870 (0.794)	1.525 (1.313)
9	0.377 (0.343)	0.872 (0.804)	1.469 (1.291)
10	0.353 (0.324)	0.874 (0.812)	1.426 (1.273)
12	0.316 (0.295)	0.878 (0.826)	1.363 (1.245)
15	0.277 (0.263)	0.885 (0.842)	1.303 (1.215)
20	0.236 (0.226)	0.893 (0.861)	1.243 (1.182)
30	0.189 (0.184)	0.906 (0.884)	1.182 (1.144)
50	0.144 (0.142)	0.922 (0.908)	1.130 (1.109)
n	$\beta = 0.9545 (\Delta L = 2)$		
	$\frac{\Delta a}{\hat{\sigma}}$	$\frac{\sigma_1}{\hat{\sigma}}$	$\frac{\sigma_2}{\hat{\sigma}}$
2	13.968 (2.528)	0.621 (0.471)	49.59 (4.366)
3	3.201 (1.671)	0.630 (0.526)	8.074 (3.042)
4	1.909 (1.311)	0.647 (0.564)	4.449 (2.511)
5	1.435 (1.107)	0.663 (0.593)	3.296 (2.223)
6	1.185 (0.974)	0.678 (0.616)	2.744 (2.041)
7	1.027 (0.878)	0.690 (0.635)	2.423 (1.914)
8	0.918 (0.805)	0.701 (0.651)	2.212 (1.821)
9	0.837 (0.748)	0.711 (0.665)	2.062 (1.748)
10	0.773 (0.701)	0.720 (0.677)	1.951 (1.691)
12	0.680 (0.629)	0.735 (0.698)	1.795 (1.604)
15	0.587 (0.553)	0.753 (0.722)	1.649 (1.515)
20	0.491 (0.471)	0.776 (0.751)	1.511 (1.424)
30	0.388 (0.378)	0.806 (0.788)	1.376 (1.326)
50	0.293 (0.289)	0.841 (0.829)	1.265 (1.239)

Таблица 4: Наиболее вероятные и средние значения некоторых оценок параметра σ Гауссова распределения по n измерениям x_i .

Величина	$\frac{\hat{\sigma}}{\sigma}$	$\frac{\tilde{\sigma}}{\sigma}$	$\left(\frac{\hat{\sigma}}{\sigma}\right)^2$	$\left(\frac{\tilde{\sigma}}{\sigma}\right)^2$
Наиболее вероятное	$\sqrt{\frac{n-2}{n}}$	$\sqrt{\frac{n-2}{n-1}}$	$\frac{n-3}{n}$	$\frac{n-3}{n-1}$
Среднее значение	$\frac{(n-2)!!}{(n-3)!!} \cdot \sqrt{\frac{R_n}{n}} a$	$\frac{(n-2)!!}{(n-3)!!} \cdot \sqrt{\frac{R_n}{n-1}}$	$\frac{n-1}{n}$	1

^aЗдесь использованы обозначения:

$$n!! = n \cdot (n-2) \cdot (n-4) \cdots, \quad R_{2k} = 2/\pi, \quad R_{2k+1} = \pi/2.$$

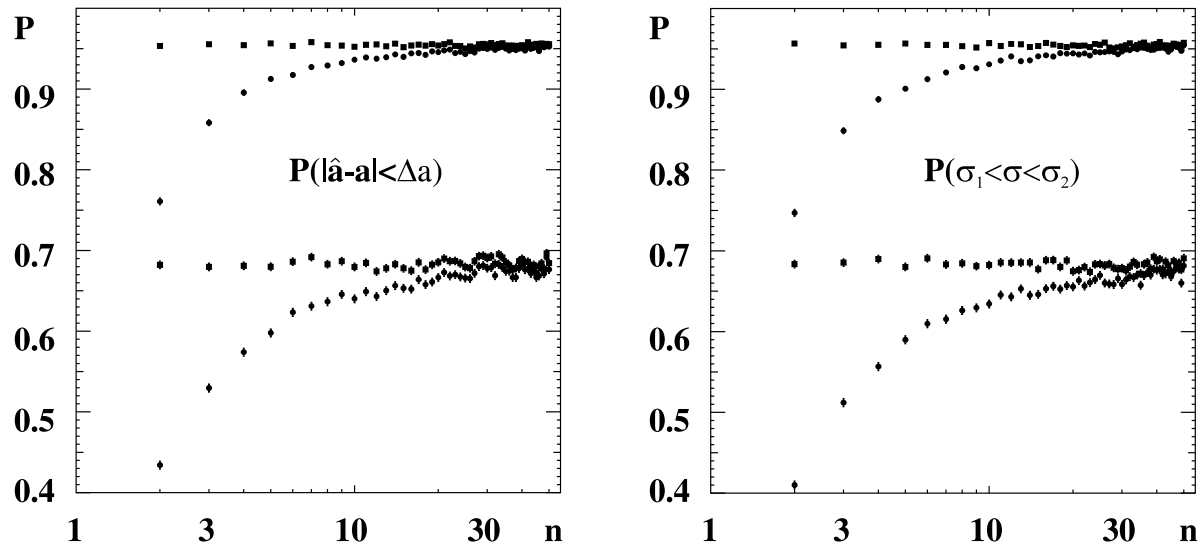


Рис. 3: Вероятность P покрыть истинное значение параметров a (слева) и σ (справа) Гауссова распределения доверительным интервалом для двух уровней доверия: 68.27% и 95.45%. \bullet — доверительный интервал определяется по функции правдоподобия, \blacksquare — доверительный интервал определяется методом Неймана. Статистика расчета в каждой точке равна 10^4 .

4. Дискретные распределения

4.1. Распределение Пуассона

Рассмотрим простейший случай, когда по единственному измерению n мы пытаемся определить среднее значение p Пуассонова распределения:

$$W_k = \frac{p^k}{k!} e^{-p}. \quad (26)$$

Функция правдоподобия в этом случае может быть записана в виде

$$L = p - n + n \cdot \ln \frac{n}{p} \quad (27)$$

(здесь использовано то обстоятельство, что к логарифмической функции правдоподобия можно добавить любое слагаемое, не зависящее от параметра p). График этой функции всегда несимметричен (пример графика при $n = 2$ приведён на рис.4). Легко получить оценку параметра p по методу максимального правдоподобия:

$$\hat{p} = n. \quad (28)$$

68%-ный доверительный интервал $(\hat{p} - \Delta p_1, \hat{p} + \Delta p_2)$ получается решением уравнения

$$\Delta p - n \cdot \ln \left(1 + \frac{\Delta p}{n} \right) = \frac{1}{2} \quad (29)$$

относительно Δp . Очевидно, что при любом $n > 0$ существует ровно два решения в области $\Delta p > -n$. В пределе больших n величина $\frac{\Delta p}{n}$ становится малой и логарифм можно разложить в ряд до второго члена, тогда решение получается в следующем виде

$$\frac{\Delta p^2}{2n} = \frac{1}{2} \quad \implies \quad \Delta p = \pm \sqrt{n} \quad (30)$$

Это отражает тот факт, что при большой статистике распределение Пуассона переходит в распределение Гаусса со средним n и среднеквадратичным отклонением \sqrt{n} . При малой статистике уравнение (29) можно решать любым численным способом, например, методом деления пополам. Можно также использовать метод последовательных приближений с рекуррентными формулами. Например, можно использовать такой подход. Если обозначить за величину $x = \Delta p/n$, то для положительной ветви

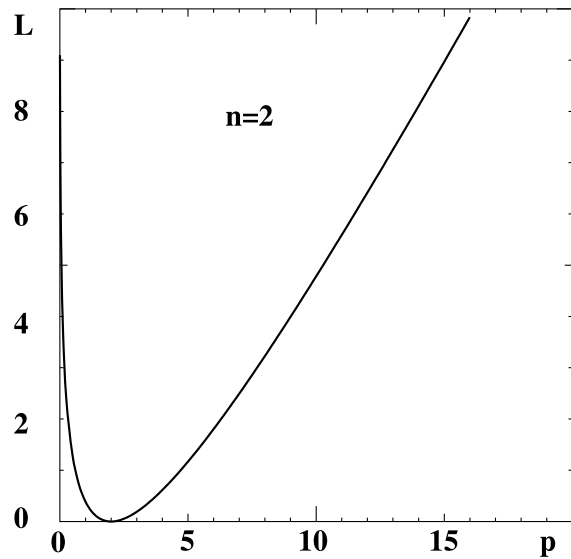


Рис. 4: Функция правдоподобия для Пуассонова распределения вероятностей.

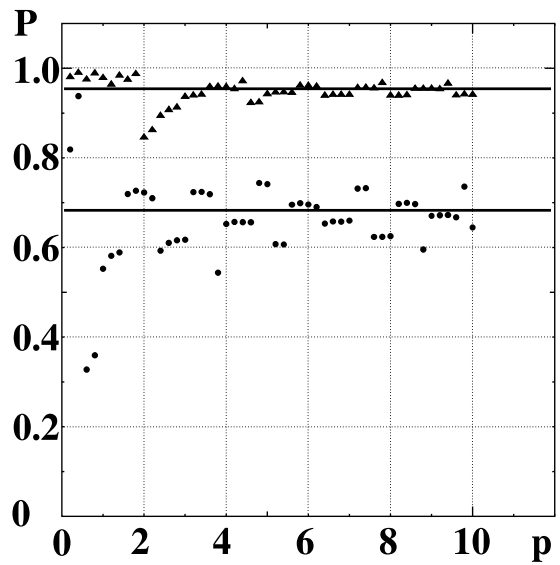


Рис. 5: Вероятность P попадания истинного значения параметра p Пуассонова распределения в доверительный интервал, полученный методом максимального правдоподобия. \bullet — 68%-ый доверительный интервал, \blacktriangle — 95%-ый доверительный интервал. Горизонтальные сплошные линии показывают “правильное” значение вероятности (0.6827 и 0.9545).

решения можно использовать формулу

$$x_{k+1} = \frac{[\ln(1+x_k) + \frac{1}{2n}] \cdot (1+x_k) - x_k}{x_k}; \quad x_1 = 1 \quad (31)$$

и для отрицательной ветви

$$x_{k+1} = \exp\left(x_k - \frac{1}{2n}\right) - 1; \quad x_1 = -1. \quad (32)$$

В случае $n = 0$ при прямолинейном использовании рекомендаций метода максимального правдоподобия, оптимальное значение $\hat{p} = 0$ и 68%-ый доверительный интервал получается $(0, 0.5)$.

Полученный таким образом доверительный интервал можно проверить методом Монте-Карло, вычислив вероятность P — в какой доле случаев доверительный интервал захватывает истинное значение p . Зависимость этой величины $P(p)$ представлена на рис. 5 (там же нарисован аналогичный график для 95%-ого интервала, где $\Delta L = 2$). Обращает внимание на себя медленная сходимость результата к асимптотическому при увеличении статистики вопреки традиционному мнению, что для Пуассонова распределения при статистике 10 или более событий все характеристики совпадают с характеристиками Гауссова распределения. Причиной негладкости представленной функции является не ограниченная статистика расчёта методом Монте-Карло, а дискретность распределения Пуассона.

Попытаемся построить доверительную зону в соответствии с методом Неймана. Дискретность распределения вероятностей не позволяет определять границы зоны аналогично формуле (14). Традиционное решение в этом случае — минимальным образом расширить доверительную зону так, чтобы всегда выполнялось неравенство $P(p) \geq \beta$. Такой подход называется консервативным. В нашем случае оценка параметра $\hat{p} = n$ и имеет то же самое распределение вероятностей Пуассона. В рамках консервативного подхода при построении доверительной зоны для любого значения p мы должны определить интервал суммирования (k_1, k_2) такой, чтобы

$$\sum_{k=k_1}^{k_2} \frac{p^k}{k!} e^{-p} \geq \beta. \quad (33)$$

Доверительная зона для случая $\beta = 0.6827$ изображена на рис.6. Так как результаты эксперимента здесь — всегда целые числа, то можно

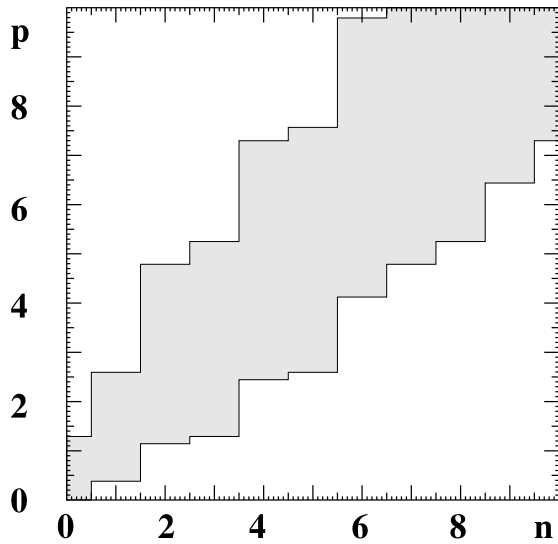


Рис. 6: Доверительная зона для оценки параметра p распределения Пуассона с уровнем доверия $\beta = 0.6827$

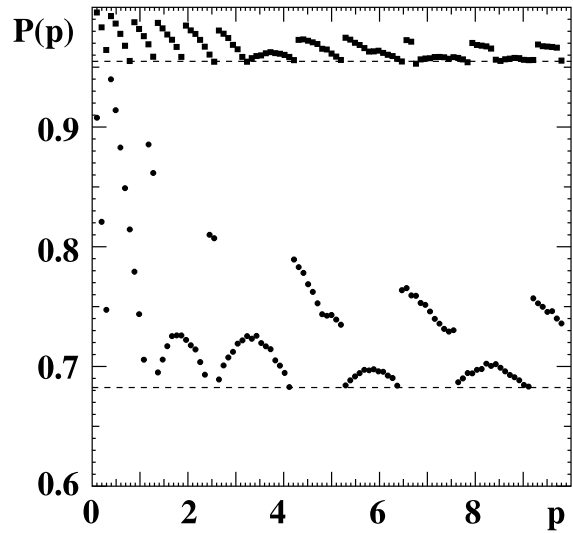


Рис. 7: Вероятность покрытия P истинного значения параметра p системой доверительных интервалов для $\beta = 0.6827$ (●) и 0.9545 (■) из табл. 5

Таблица 5: Границы доверительных интервалов с уровнем доверия β для оценки параметра p распределения Пуассона по единственному измерению n

n	$\beta = 0.6827$		$\beta = 0.9$		$\beta = 0.9545$	
	p_1	p_2	p_1	p_2	p_1	p_2
0	0.000	1.290	0.000	2.436	0.000	3.234
1	0.382	2.593	0.105	4.951	0.047	5.752
2	1.145	4.787	0.532	6.470	0.337	6.742
3	1.290	5.250	1.102	7.969	0.787	8.332
4	2.444	7.296	1.745	8.597	1.324	9.769
5	2.593	7.569	2.433	9.716	1.917	11.100
6	4.122	9.792	2.436	11.993	2.551	13.280
7	4.787	10.811	3.589	12.531	3.214	14.764
8	5.250	11.242	4.532	14.508	3.234	15.360
9	6.438	13.299	4.951	15.298	4.309	16.401
10	7.296	14.312	5.976	15.986	5.208	18.768

Таблица 6: Значения верхнего предела на параметр p распределения Пуассона с уровнем доверия $\beta = 90\%$, устанавливаемые по функции правдоподобия ($\Delta L = 0.822$)

Число событий n	0	1	2	3	4
Верхний предел	0.822	2.880	4.398	5.799	7.139

для первых нескольких значений n привести вычисленные доверительные интервалы (табл. 5). На рис. 7 показана вероятность покрытия истинного значения p системой доверительных интервалов с уровнем доверия $\beta = 0.6827$ и 0.9545 . Сравнивая вероятности покрыть истинное значение доверительным интервалом с 68%-ым уровнем доверия, полученным по функции правдоподобия (рис.5) и методом Неймана (рис.7), можно заметить, что при $p > 2$ разброс вероятности для этих методов одинаков (определяется дискретностью распределения вероятностей), однако, для метода Неймана строго выполняется условие “консервативности” $P(p) \geq 68.27\%$. Кроме того, с увеличением p разброс уменьшается, и в асимптотике оба метода будут эквивалентны.

Для малых уровней эффекта неопределенность результата возрастает, и в этом случае величину параметра p часто “оценивают сверху”.

4.2. Верхний предел при малом числе событий без вычитания фона

Рассмотрим случай, когда в результате эксперимента зарегистрировано малое число событий и уровень фона достаточно мал, чтобы им можно было пренебречь.

Оценка верхнего предела по функции правдоподобия производится по тем же формулам (27—31), что и для “двустороннего” доверительного интервала. Естественно, для получения соотношения между уровнем доверия β и превышением функции правдоподобия над минимумом ΔL надо использовать (5). Значения верхнего предела, которые устанавливаются на параметр p по функции правдоподобия с уровнем доверия 90%, приведены в табл.6. Вероятность покрытия истинного значения параметра p этой системой верхних значений приведена на рис.8. Видно, что в области малых значений параметра p вероятность покрытия истинного значения доверительным интервалом опускается до 0.6, что не может быть удовлетворительным. Посмотрим, что дает метод Неймана.

“Консервативную” доверительную зону по методу Неймана с односто-

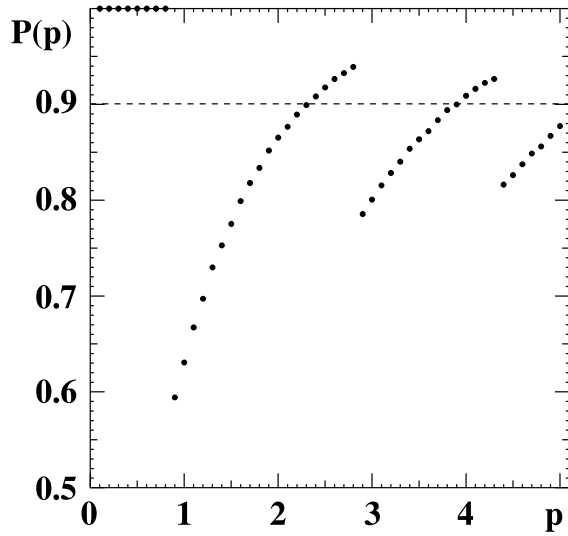


Рис. 8: Вероятность того, что истинное значение параметра p окажется ниже верхнего предела, установленного по функции правдоподобия

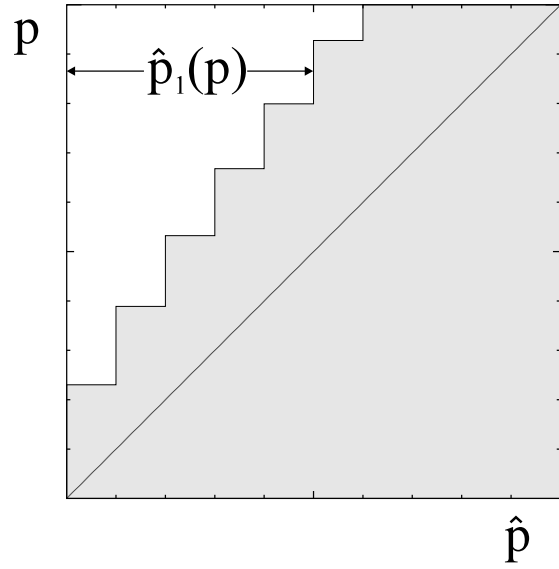


Рис. 9: Доверительная зона с уровнем доверия 90% для оценки параметра распределения Пуассона \hat{p} сверху

ронней границей можно определить следующим образом:

$$\sum_{n=\hat{p}_1(p)}^{\infty} W_n \geq \beta > \sum_{n=1+\hat{p}_1(p)}^{\infty} W_n, \quad (34)$$

где β — заданный уровень доверия. Построенная по этому правилу доверительная зона для 90%-ого уровня доверия изображена на рис.9. В табл.7 приведена величина верхнего предела на параметр Пуассонова распределения p для нескольких значений числа зарегистрированных событий. Именно так рекомендовалось экспериментаторам устанавливать верхний предел в разделе статистики в [6] до 1996 года. На рис.10 представлена вероятность покрытия истинного значения верхним пределом, устанавливаемым по этим правилам.

В работе [7] предлагается способ борьбы с рядом проблем, возникающих при построении доверительных интервалов (например, доверительные интервалы нулевой длины). В работе предложено формировать доверительную зону по горизонтали уже с учетом физических ограничений на параметры модели, а также выбирать границы зоны не по равенству вероятности выхода за правую и левую границы, а по равенству функций правдоподобия на границах. В отношении установления верхнего предела при малом числе событий в этой работе также предложено изменить

Таблица 7: Верхний предел на параметр p распределения Пуассона на 90, 95 и 99%-ном уровне доверия.

Число зарегистр. событий n	0	1	2	3	4	5
Верхний предел на параметр p ($\beta = 90\%$)	2.30	3.89	5.32	6.68	7.99	9.27
Верхний предел на параметр p ($\beta = 95\%$)	3.00	4.74	6.30	7.75	9.15	10.51
Верхний предел на параметр p ($\beta = 99\%$)	4.61	6.64	8.41	10.05	11.60	13.11

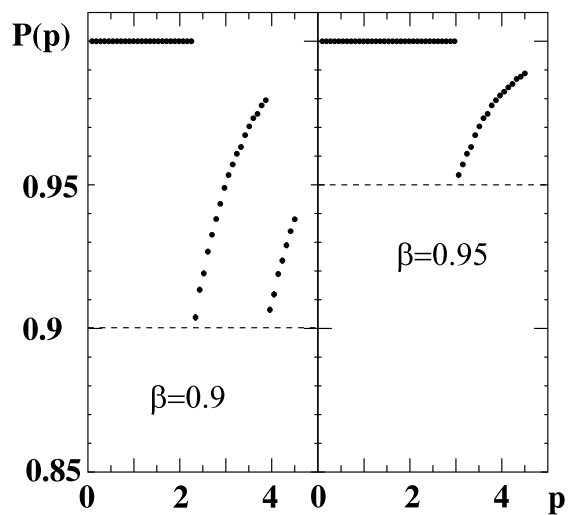


Рис. 10: Вероятность того, что истинное значение параметра p окажется ниже верхнего предела, установленного по таблице 7

Таблица 8: Таблица доверительных интервалов из [8] для параметра p Пуассонова распределения

n	0	1	2	3	4	5
$\beta = 0.9$	2.44	4.36	5.91	7.42	8.60	9.99
	0.00	0.11	0.53	1.10	1.47	1.84
$\beta = 0.95$	3.09	5.14	6.72	8.25	9.76	11.26
	0.00	0.05	0.36	0.82	1.37	1.84
n	6	7	8	9	10	11
$\beta = 0.9$	11.47	12.53	13.99	15.30	16.50	
	2.21	3.56	3.96	4.36	5.50	
$\beta = 0.95$	12.75	13.81	15.29	16.77	17.82	
	2.21	2.58	2.94	4.36	4.75	

подход: в реальном эксперименте при нулевом числе зарегистрированных событий экспериментаторы обычно устанавливают верхний предел, а при ненулевом числе событий — указывают двусторонний доверительный интервал. В соответствии с этими идеями авторами была предложена таблица 8, которая с тех пор используется в [8] в качестве рекомендации для установки верхних пределов. Проверка методом Монте-Карло зависимости $P(p)$ приведена на рис.11.

Теперь для случая, когда стратегия экспериментатора меняется в зависимости от результата (для нуля событий — верхний предел, в остальных случаях — двусторонний доверительный интервал), попробуем минимизировать величину превышения вероятности $P(p)$ над уровнем β . Для заданного уровня доверия β эта задача математически соответствует следующей формулировке: для любого значения параметра p подобрать интервал суммирования (k_1, k_2) для функции

$$P_{k_1 k_2}(p) = \sum_{k=k_1}^{k_2} \frac{p^k}{k!} e^{-p}, \quad 0 \leq k_1 \leq k_2 < \infty \quad (35)$$

так, чтобы всегда выполнялось неравенство $P_{k_1 k_2}(p) - \beta \geq 0$, но величина превышения была минимальной. При малых p такую минимизацию можно провести численно и составить таблицу доверительных интервалов (табл.9). Вероятность покрытия истинного значения параметра p доверительными интервалами из этой таблицы представлена на рис.12.

Сравнивая все рассмотренные варианты вычисления верхнего предела при малом числе зарегистрированных событий, можно сказать, что

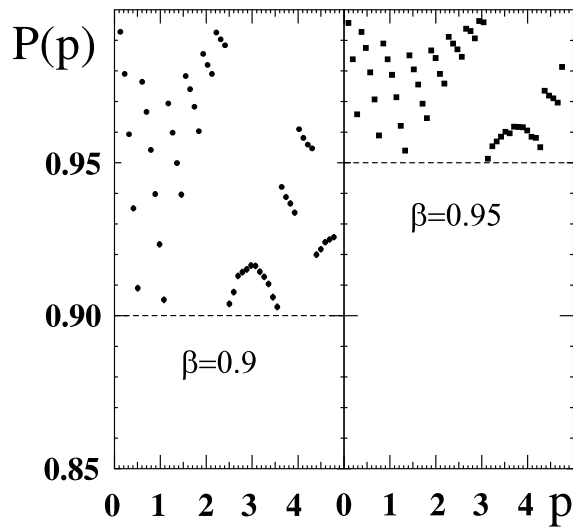


Рис. 11: Вероятность $P(p)$ покрытия истинного значения параметра p доверительным интервалом из табл.8 с уровнем доверия β (\bullet — $\beta = 0.9$, \blacksquare — $\beta = 0.95$)

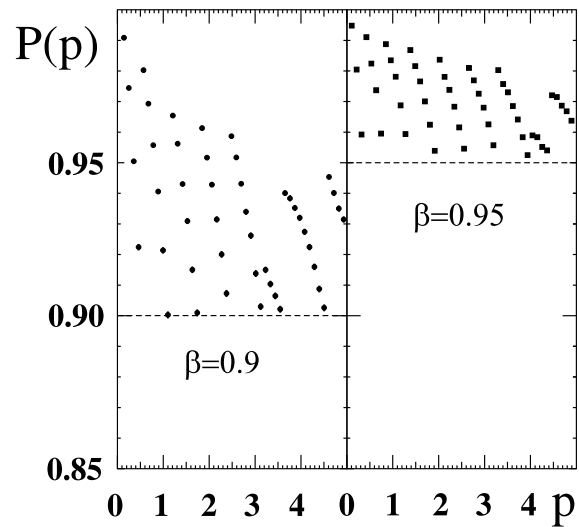


Рис. 12: Вероятность $P(p)$ “покрытия” истинного значения параметра p доверительным интервалом с уровнем доверия β (\bullet — $\beta = 0.9$, \blacksquare — $\beta = 0.95$). Доверительная зона построена численной минимизацией превышения вероятности “покрытия” над уровнем доверия

Таблица 9: Таблица доверительных интервалов для Пуассона распределения, полученная численной минимизацией превышения над уровнем доверия β вероятности покрытия истинного значения параметра p доверительным интервалом

n	0	1	2	3	4	5
$\beta = 0.9$	3.15	5.38	6.91	8.50	9.13	10.81
	0.00	0.11	0.53	1.10	1.75	2.43
$\beta = 0.95$	3.98	6.12	7.58	9.08	10.63	12.23
	0.00	0.05	0.36	0.82	1.37	1.97
n	6	7	8	9	10	11
$\beta = 0.9$	12.53	13.16	14.97	15.58	17.47	18.07
	3.15	3.59	4.53	5.38	5.98	6.91
$\beta = 0.95$	13.86	14.50	16.20	17.91	18.57	20.35
	2.61	3.29	3.98	4.46	5.32	6.12

приближенный метод построения доверительного интервала по функции правдоподобия дает неудовлетворительный результат в случаях $n = 0, 1, 2$. Из остальных вариантов предпочтительным все же представляется табл. 7. Рассуждение, что в зависимости от полученного результата у экспериментатора меняется стратегия, здесь представляется неосновательным. Если уж пытаться следовать практике представления экспериментальных результатов, то тогда следовало бы учесть, что при нулевом числе событий устанавливается верхний предел на уровне доверия 90%, а при ненулевом числе событий — двусторонний доверительный интервал с уровнем доверия 68%. Такая стратегия с переменной уровня доверия в зависимости от результата не укладывается в рамки классического метода построения доверительной зоны. Кроме того, в зависимости от задачи может потребоваться установить верхний предел и при ненулевом количестве событий. Кстати, из всех “консервативных” вариантов табл. 7 дает наиболее жесткое ограничение при нулевом числе событий на уровне доверия 90% — $p < 2.3$.

В случае, когда требуется вычитать расчетный фон, необходимо строить доверительную зону для конкретного эксперимента. Очевидно, что это необходимо делать, когда число событий меньше или порядка пяти. Варианты таких построений подробно описаны в [7]. Здесь мы еще рассмотрим одну иллюстрацию к дискретным распределениям вероятности.

4.3. Подгонка резонансной кривой

В экспериментах на коллайдерах, когда зависимость числа событий от энергии в системе центра масс подгоняется кривой возбуждения резонанса с целью получить параметры изучаемого резонанса (массу, ширину и т.п.), распределение вероятностей по полученным оценкам является дискретным, хотя это распределение часто бывает достаточно плотным. Построение доверительной зоны для таких случаев представляется достаточно трудоемкой задачей, и результат будет зависеть от многих обстоятельств: распределения интеграла светимости по энергии, наличия фона, теоретической модели и т.п. Очевидно, что этим надо заниматься в тех случаях, когда есть уверенность, что построение доверительных интервалов по функции правдоподобия дает неверный результат. Но даже эта более простая задача проверки доверительных интервалов может потребовать большой вычислительной мощности, так как при обработке экспериментальных данных обычно проводится численная минимизация логарифмической функции правдоподобия, и численное моделирование многократного повторения обработки эксперимента при разных истин-

ных значениях параметров теоретической модели может представлять сложную техническую задачу.

Здесь мы рассмотрим совсем упрощенную модель эксперимента по определению параметров резонанса.

Пусть в десяти точках по энергии получено число событий какого-либо канала распада резонанса, пусть также нет никакого фона (что редко бывает) и свободных параметров модели всего два: масса резонанса и интенсивность распада. То есть видимая форма резонанса фиксирована и определяется, например, известным аппаратурным разрешением σ , а истинная ширина резонанса много меньше этой величины. Для простоты выберем эту форму в виде Гауссова распределения, хотя обычно даже в таких простых случаях она имеет несимметричный вид.

Итак, проведен эксперимент, в котором в точках с энергией

$$E_i = E_0 + (i - 5.5) \cdot \sigma/2, \quad i = 1, 2, \dots, 10$$

с равным интегралом светимости измерено число событий n_i распада резонанса по изучаемому каналу. Ожидаемое число событий в точке i равно

$$p_i = B \cdot \exp \left[-\frac{(m - E_i)^2}{2\sigma^2} \right]. \quad (36)$$

Функция правдоподобия, зависящая от параметров m и B , может быть записана в виде

$$L = \sum_{i=1}^{10} \left[p_i - n_i + n_i \cdot \ln \frac{n_i}{p_i} \right]. \quad (37)$$

В случае $n_i = 0$ последнее слагаемое в скобках полагается равным нулю. По параметру B минимизация может быть проведена аналитически:

$$B_0 = \frac{\sum_{i=1}^{10} n_i}{\sum_{i=1}^{10} \exp \left[-\frac{(m - E_i)^2}{2\sigma^2} \right]}, \quad (38)$$

а минимизация по параметру m может быть быстро выполнена, например, методом “золотого сечения” [9].

Возьмем теперь истинное значение $m = E_0$, и в зависимости от истинного значения B проверим вероятность покрытия истинных значений

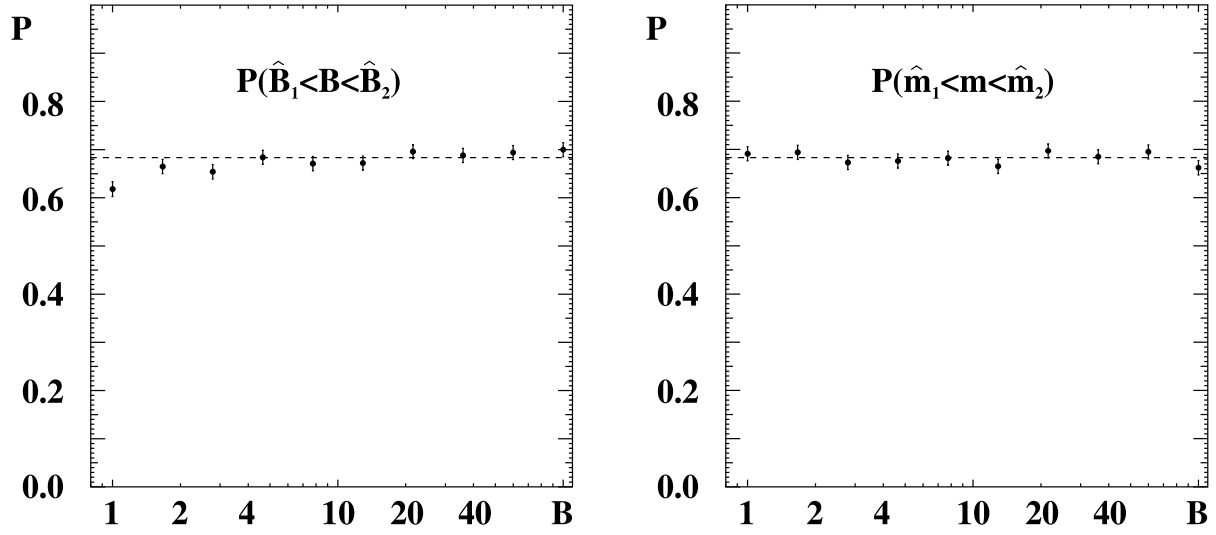


Рис. 13: Вероятность покрытия истинных значений параметров B и m резонансной кривой доверительными интервалами (\hat{B}_1, \hat{B}_2) и (\hat{m}_1, \hat{m}_2) , соответственно, полученными по функции правдоподобия на уровне $\Delta L = 1/2$ (ожидаемый уровень доверия 68%). Статистика расчетов методом Монте-Карло в каждой точке — 10^3 “экспериментов”

m и B 68%-ми доверительными интервалами, полученными по функции правдоподобия. Легко заметить, что среднее значение

$$\langle \sum_{i=1}^{10} n_i \rangle = B \cdot \sum_{i=1}^{10} \exp \left[-\frac{(i - 5.5)^2}{8} \right] \approx 5 \cdot B. \quad (39)$$

Поэтому интервал B от 1 до 100 соответствует средней статистике эксперимента от 5 до 500. На рис.13 приведена вероятность покрытия истинных значений B и m доверительными интервалами, полученными по функции правдоподобия на уровне $\Delta L = 1/2$ (68%-ый доверительный интервал).

Результат получился несколько неожиданный — даже при статистике порядка 5 событий оценка доверительного интервала по функции правдоподобия дает достаточно точную вероятность покрытия истинного значения. Для оценки массы резонанса даже не видно ухудшения при низкой статистике порядка 5 событий. Возможно, это объясняется отсутствием сильной корреляции между параметрами B и m (коэффициент линейной корреляции в нашем случае равен нулю). По-видимому, если бы свободными параметрами модели были полная ширина на полувысоте Γ и интенсивность распада B , то точность оценки обоих параметров ухудшалась с уменьшением статистики.

5. Заключение

В работе для нескольких примеров, взятых из экспериментальной физики элементарных частиц, проведен расчет методом Монте-Карло вероятности покрытия истинных значений параметров теоретической модели доверительными интервалами, оцененными по логарифмической функции правдоподобия. При малой статистике, обычно, появляется заметное отклонение этой величины от ожидаемого уровня доверия.

В простых случаях вычислены доверительные интервалы по классическому методу Неймана, которые обеспечивают точное соответствие вероятности покрытия истинного значения этими интервалами заявленному уровню доверия. Полученные точные значения доверительных интервалов могут быть использованы при обработке соответствующих экспериментов (например, измерение времени жизни при небольшом числе зарегистрированных распадов).

В сложных случаях, когда функция правдоподобия зависит от многих параметров и построение доверительной зоны по методу Неймана затруднительно, а статистика эксперимента недостаточно велика, предлагается проверять методом Монте-Карло соответствие вероятности покрытия доверительными интервалами истинного значения параметров заявленному уровню доверия.

Проведено также сравнение вероятности покрытия истинного значения параметра p распределения Пуассона доверительными интервалами на 90%-ом уровне доверия, предлагаемыми в статистическом разделе современных таблиц свойств частиц [8], с рекомендуемыми верхними пределами в этих же таблицах до 1998 г. В результате сравнения представляется необоснованным переход в 1998 г. от верхнего предела $p < 2.30$ к $p < 2.44$ при нулевом количестве зарегистрированных событий (хотя разница и невелика).

Литература

- [1] *В.Идье, Д.Драйард, Ф.Джеймс, М.Рус, Б.Садуле.* Статистические методы в экспериментальной физике. Пер. с англ. В.С.Курбатова. Под ред. А.А.Тяпкина. М.: Атомиздат, 1976.
- [2] *Яноши Л.* Теория и практика обработки результатов измерений. М.: Мир, 1965.
Janossy, L. Theory and Practice of the Evaluation of Measurements. Clarendon Press, 1965.
- [3] *Худсон Д.* Статистика для физиков. М.: Мир, 1970.
Hudson, Derek J. Lectures on Elementary Statistics and Probability. CERN, 1963 (CERN 63-29).
- [4] *Maurice G.Kendall and Alan Stuart.* The advanced theory of statistics. 4th ed., Macmillan, 1977
М.Кендалл, А.Стьюарт. Статистические выводы и связи. Пер. с англ. Л.И.Гальчука, А.Т.Терехина, под редакцией А.Н.Колмогорова. М.: "Наука", 1973.
- [5] *Neyman J.* Trans. Royal Soc. London, Ser. A, **236** (1937) 333-380. Reprinted in "A Selection of Early Statistical Papers on J.Neyman" (University of California Press, Berkeley, 1967), pp. 250-289.
- [6] *Particle Data Group.* Review of particle physics. Phys. Rev. D54 (1996) 1-720.
- [7] *Feldman G.J., Cousins R.D.* A Unified Approach to the Classical Analysis of Small Signals. Phys. Rev. D57 (1998) 3873-3889.
e-Print archive: physics/9711021.
- [8] *Particle Data Group.* Review of particle physics. Eur. Phys. J. C15: 1, 2000.
- [9] *James F.* Function minimization // Proceedings of the 1972 CERN computing and data processing school. Pertisau, Austria, 10-24 September, 1972.
Preprint CERN 72-21, 1 Dec. 1972, pp. 1-52.