Современные методы математической статистики

(Два подхода к вероятности)

А. Нозик

В прошлой серии…

To be continued…

- Принципиальное отличие частотного и Байесовского подхода: что выбрать?
- Использование функций значимости для планирования эксперимента
- Программная реализация всего вышеописанного.

Как интерпретировать результаты?

$$A = -2.51 \ \pm 0.3_{stat} \pm 0.2_{syst}$$

Что обозначают эти цифры?

Формальное определение вероятности

Частотная вероятность

Байесовская вероятность

Обобщенный подход?

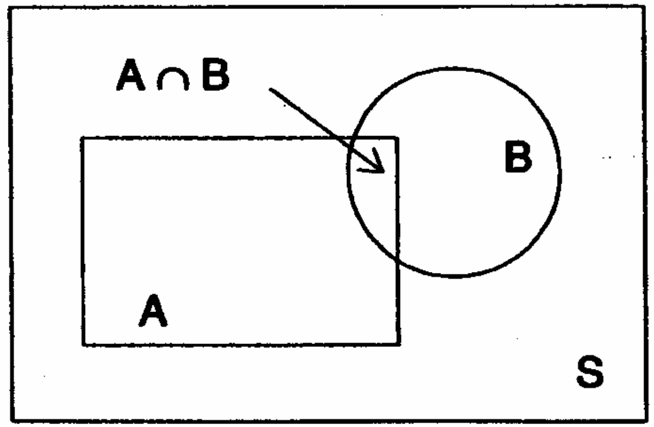
Формальное определение вероятности (Колмогоров –1933)

Рассмотрим пространство S, называемый выборочным пространством или пространством образцов (sample space) , содержащий определенный набор элементов, конкретизацию которых пока оставим открытой. Каждому поднабору (subset) $A \in S$ приписывается действительное число $P(A)$, называемое вероятностью и определяемой следующими аксиомами:

1. Для каждого A из S $P(A) \geq 0$
2. Для любых двух A и B, которые *несовместны*, т.е. взаимно исключены $(A \cap B = \varnothing)$, вероятность, приписываемая объединению A и B, есть сумма двух соответствующих вероятностей: $P(A \cup B) = P(A) + P(B)$
3. Вероятность, приписываемая всему пространству образцов, есть единица; $P(S) = 1$.

Теорема Байеса (условная вероятноть)

Условная вероятность : $P(A \mid B) = \dfrac{P(A \bigcap B)}{P(B)}$



Верно и симметричное утверждение: $P(B \mid A) = \dfrac{P(B \bigcap A)}{P(A)}$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Вероятность как относительная частота

- Пусть S – множество всевозможных исходов какого-либо эксперимента.
- Пусть A – некоторое подмножество S.
- Допустим, что есть возможность повторять эксперимент неограниченное количество раз. Количество исходов эксперимента, когда результат попал в A обозначим N(A), а полное количество экспериментов N(S).

В таком случае вероятностью P(A) назовем предел частоты исходов из A при стремлении количества экспериментов к бесконечности:

$$\lim_{N(S)\to\infty} \frac{N(A)}{N(S)}$$

Сложности практического применения частотного подхода

- Определение строго говоря имеет смысл только при стремлении количества измерений к бесконечности.
- Нет возможности интерпретировать уникальные события (падение метеорита, взрыв сверхновой).
- Возникает сложность с интерпретацией измерения детерминированых величин и систематических ошибок.

Напирмер:

Вероятность для неизвестной константы не имеет смысла в терминах предельной частотной интерпретации, т.к., если повторять эксперимент, зависящий от физического параметра, чья точная величина не определена (например, масса частицы), ее значение или никогда или всегда находится в фиксированном интервале.

Субъективная (Байесовская) вероятность

- Пусть S – пространство всевозможных гипотез (пространство исходов эксперимента – частный случай).
- Элементарные гипотезы взаимно исключены.
- Подмножество, содержащее более одной гипотезы истинно, если хоть какая-либо из гипотез в подмножестве истинна.
- Одна из гипотез должна быть с необходимостью истинной.

В таком случае субъективной вероятностью P(A) назовем степень веры в то, что гипотеза A верна.

В качестве выбора критерия для степени веры можно взять например относительную частоту.

Теорема Байеса и субъективная вероятность

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \longrightarrow P_{post} = \frac{P \cdot \pi_{prior}}{\int \pi_{prior}}$$

- Априорная вероятность π – степень ожидания того или иного результата до проведения эксперимента. Может включать в себя обобщенный результат предыдущих экспериментов.
- Постериорная вероятность P_{post} – степень доверия к тому или иному результату эксперимента с учетом априорной вероятности.

Пример использования априорной вероятности: шулерский кубик.

Недостатки субъективной вероятности

Она субъективная!

Тем не менее...

- Во всей современной экспериментальной физике явно или не явно используется субъективная вероятность (даже в тех случаях, когда априорной информации нет и теорема Байеса не используется).
- В случае явного согласования критериев, по которым определяется субъективная вероятность, субъективное определение открывает большое количество возможностей.

$$A = -2.51 \pm 0.3_{stat} \pm 0.2_{syst}$$

Результаты экспериментов интерпретируются только в рамках субъективного определения вероятности

Почему не удобно работать в частотной интерпретации

$$A = -2.51 \pm 0.3_{stat} \pm 0.2_{syst}$$

- Частнотная интерпретация вероятности не имеет смысла при однократном проведвении эксперимента.
- Истинное значение измеряемой величины не является случайным и не может подчиняться статистическому распределению с конечной шириной.
- Систематическую ошибку можно частотно определеить только в очень редких случаях.
- В частотной интерпретации нельзя складывать результаты различных измерений.

Как же быть?

Разумный компромисс:

Субъективная интерпретация вероятности, где в качестве субъективной вероятности берется частота выпадения результата в эксперименте.

- Детерминированные процессы с неполной информацией могут трактоваться как случайные.
- Результаты одни и те же для разных наблюдателей, использующих одни и те же данные.
- Сохраняет все полезные свойства Байесовской вероятности (сложение результатов, систематические ошибки и. т. д.). При необходимости, в анализ в любой момент может быть добавлена дополнительная априорная вероятность.

Практическое использование концепции субъективной вероятности

- Интерпретация систематических ошибок
- Сложение результатов независимых экспериментов
- Учет априорной информации о поведении исследуемого параметра (физические ограничения)
  - Проблема замены переменных и априорная вероятность Джеффри
  - Другие способы учета физической границы

Интерпретация систематических ошибок как недостающей информации

$$L(\theta,\varphi) \longrightarrow L_{post}(\boldsymbol{\theta}) = N\int L\,(\boldsymbol{\theta},\boldsymbol{\varphi})\pi(\boldsymbol{\varphi})d\boldsymbol{\varphi}.$$

- Пусть есть некоторый «систематический» параметр φ, значение которого известно с ограниченной точностью.
- В субъективной интерпретации распределение может возникать не только как частота случайного процесса, но и как мера недостаточности информации о параметре.
  Строго известный природе параметр может иметь распределение!
- В этом случае априорная информация о систематическом параметре – это просто мера недостаточности информации о его истинном значении.

Определение систематической ошибки в частотной интерпретации весьма проблематично. Оно вообще имеет смысл только когда систематическая ошибка действительно отражает истинно случайный процесс, а не сбой калибровки.

Сложение результатов различных экспериментов

$$L(\theta)_{post} = L_1(\theta) \cdot L_2(\theta)$$

- При сложении результатов различных экспериментов правдоподобие одного из них можно считать априорной вероятностью для вычисления апостериорной вероятности другого.

В частотной интепретации нельзя складывать результаты различных экспериментов, поскольку они отражают частоты различных процессов.

Априорная информация о параметре

$$L_{post}(\boldsymbol{\theta}) = N \int L\,(\boldsymbol{\theta}, \boldsymbol{\varphi}) \pi(\boldsymbol{\theta}, \boldsymbol{\varphi}) d\boldsymbol{\varphi}.$$

- Априорная информация может описывать не только знание о систематических параметрах и предыдущих экспериментах, но и знание о физических границах исследуемого параметра.

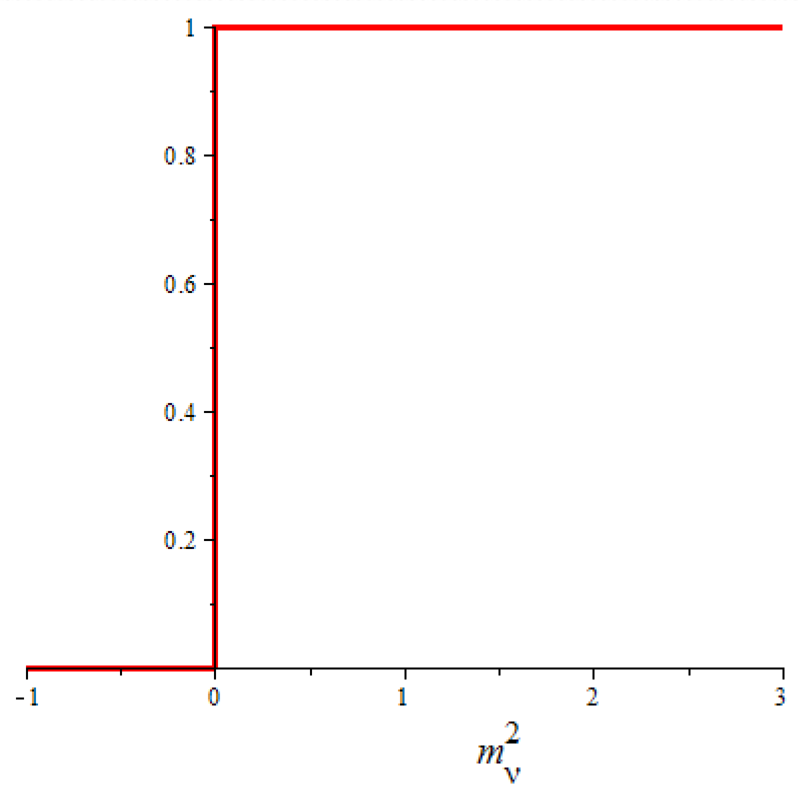
Например:

$$m_\nu^2 \geq 0$$
$$0 \leq U^2 \leq 1$$

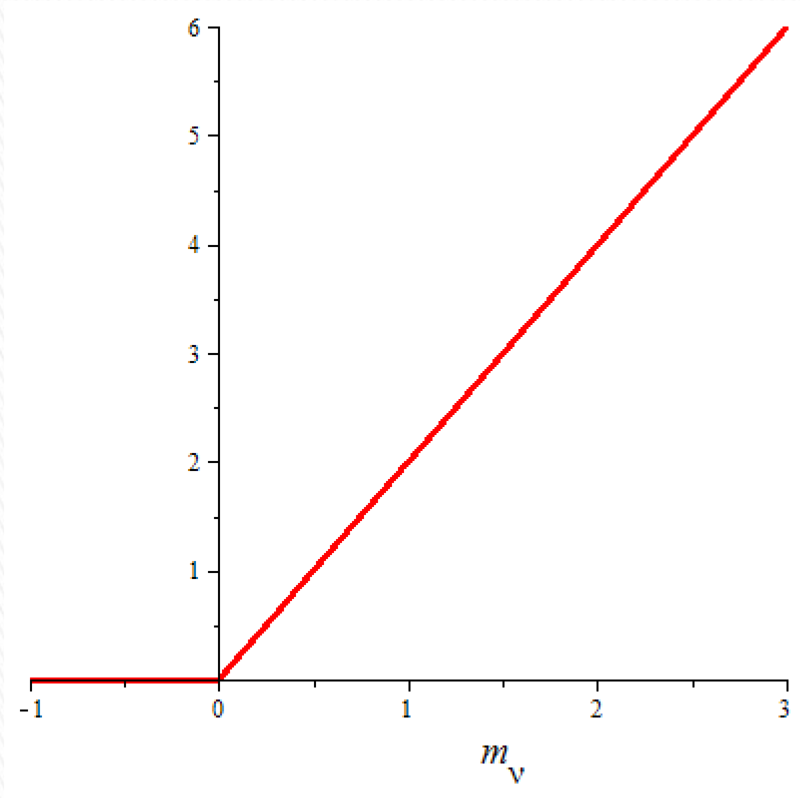
В этом случае априорная информация принимает например такой вид (простое ограничение):

$$\pi(\theta, \boldsymbol{\varphi}) = \pi(\boldsymbol{\varphi}) * \begin{cases} 0, & m_\nu^2 < 0 \\ 1, & m_\nu^2 \geq 0 \end{cases}$$

Проблема замены переменных



Квадрат массы



Масса

Априорная информация Джеффри

- Априорная информация, не меняющаяся при замене переменных (Proc. R. Soc. Lond. A 24 September 1946 **vol. 186 no. 1007** 453-461):

$$\pi_J(\theta) = \sqrt{I(\theta)}$$

$$I(\theta) = E\left[\left(\frac{d\ln L}{d\theta}\right)^2\right]$$

Для нормального распределения с фиксированной шириной

$$\pi_J(\theta) = const$$

При определении параметра с квази-гауссовским распределением и ошибкой, не зависящей от значения «простое ограничение» является инвариантным по отношению к замене координат

Другие способы учета физической границы

- Объединенный подход Фельдмана и Казинса.
- Предел чувствительности

Оба подхода очень просты в использовании, но используются только для «простых ограничений» и только в одномерном случае.

Метод Фельдмана и Казинса вызывал много дополнительной критики.

Современные возможности

$$L_{post}(\boldsymbol{\theta}) = N \int L\left(\boldsymbol{\theta}, \boldsymbol{\varphi}\right) \pi(\boldsymbol{\theta}, \boldsymbol{\varphi}) d\boldsymbol{\varphi}.$$

- Этот интеграл содержит всю информацию, которая нужна для анализа (статистика, систематика, физические границы).
- В многомерном случае его вычисление – утомительная процедура (требуется гораздо больше вычислений чем для обычного фита).
- Для этого человек изобрел компьютер!

Глобальный вывод

- Оба подхода (и частотный и субъективный) абсолютно корректны с точки зрения формальной математики.
- Существует глобальная неопределенность в выборе определения вероятности: в ортодоксальном частотном варианте четко формулируется задача, но нет возможности «честно» интепретировать результаты. В ортодоксальном субъективном определении, результаты интерпретируются, но постановка задачи является субъективной.
- В рамках чисто частотного подхода интерпретация результатов эксперимента в большинстве случаев просто невозможна.
- В рамках абсолютно субъективного подхода нет возможности универсально сформулировать задачу.
- Разумный компромис между двумя определениями дает возможнсть с одной стороны использовать в полной мере инструментарий субъективной вероятности, с другой четко формулировать задачи.
- Современная вычислительная техника позволяет избегать упрощений, и приближений, которых невозможно было избежать 10-15 лет назад.

Дополнительные слайды

Функция правдоподобия

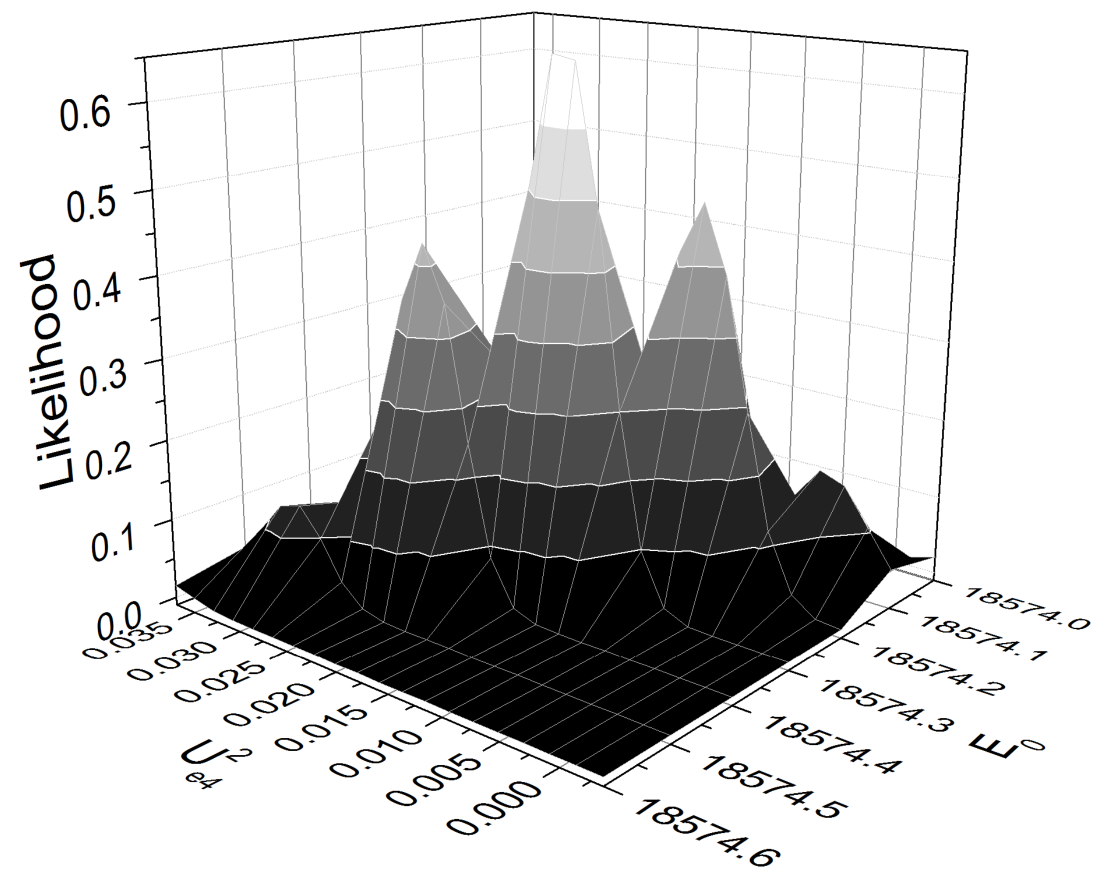
$$L(\boldsymbol{D}|\boldsymbol{\theta}) = \prod_i P(D_i, \boldsymbol{\theta})$$

$$\ln(L(\boldsymbol{D}|\boldsymbol{\theta})) = \sum_i \ln(P(D_i, \boldsymbol{\theta}))$$

Оценка θ: $L \rightarrow max$ является состоятельной, несмещенной и в большинстве случаев обладает максимальной возможной эффективностью (граница Рао-Крамера).

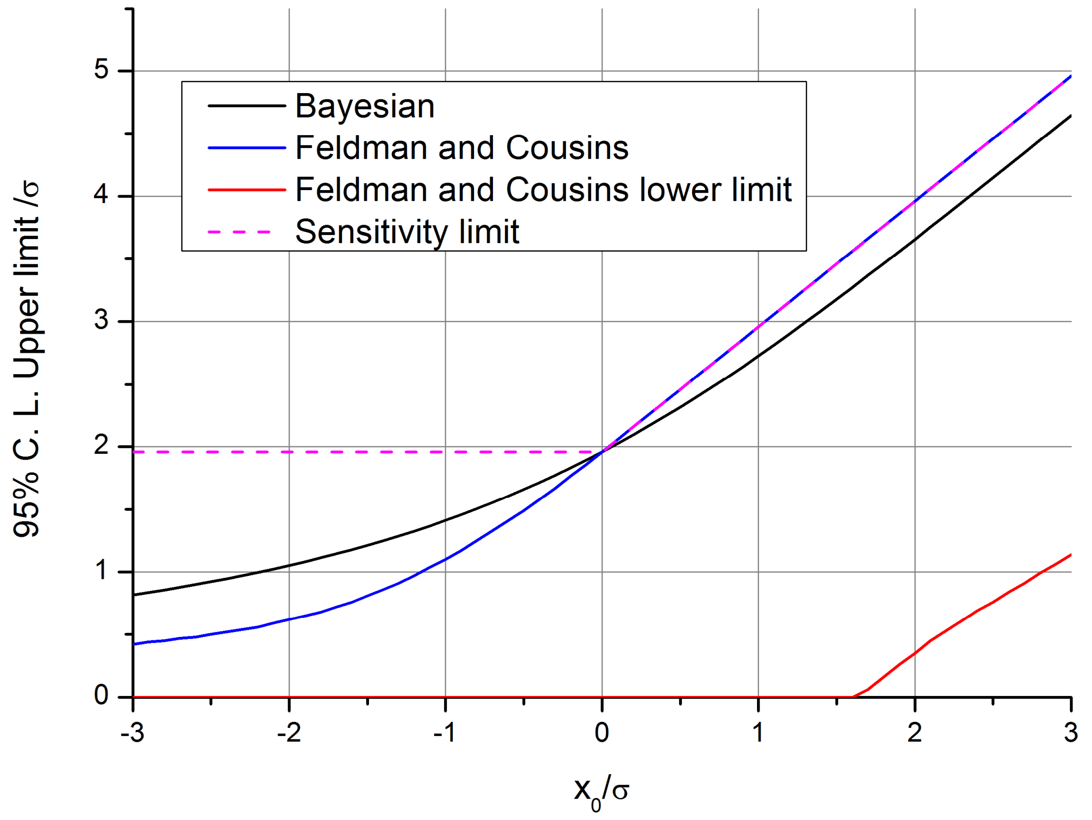
Для гауссовского распределения $P(D_i, \boldsymbol{\theta}) \sim -\frac{\left(X_i/T_i - \mu(E_i, \boldsymbol{\theta})\right)^2}{2\sigma_i^2}$,

$$\ln(L) \sim -\frac{\chi^2}{2}$$



Пример негауссовскоо распределения

	Достоинства	Недостатки
Unified approach of Feldman and Cousins	• Очень просто использовать	• Требуется продолжение модели в нефизическую область. • Обоснованность метода под большим вопросом • Интервалы не центральные • При «отицательном» значении среднего получается сильно заниженная оценка верхней границы
Байесовский подход	• Гибкий, хорошо обоснованный подход. • Можно оценивать верхние и нижние границы (одновременно) и использовать априорную информацию любого вида. • Можно легко складывать результаты различных экспериментов.	• Требуется явное использование априорной (субъективной) информации.
Sensitivity limit	• Нет зависимоить верхнего предела от «отрицательности» оценки.	• Вся информация об «отрицательности» параметра игнорируется. • Пока нет строгого обосноания (ждем статью А. Лохова и Ф. Ткачева)



Сравнение разных методов учета границы