3-681

4947/4-77

ОБЪЕДИНЕННЫЙ
ИНСТИТУТ
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ

ДУБНА

ЛЯП

Р10 - 10855

В.Б.Злоказов

ПРОГРАММА ДЛЯ ПОСТРОЕНИЯ ОЦЕНОК ПАРАМЕТРА ЭКСПОНЕНЦИАЛЬНОГО РАСПРЕДЕЛЕНИЯ В УСЛОВИЯХ МАЛОЙ СТАТИСТИКИ

1977

Р10 - 10855

В.Б.Злоказов

ПРОГРАММА ДЛЯ ПОСТРОЕНИЯ ОЦЕНОК ПАРАМЕТРА ЭКСПОНЕНЦИАЛЬНОГО РАСПРЕДЕЛЕНИЯ В УСЛОВИЯХ МАЛОЙ СТАТИСТИКИ

Направлено в "Nuclear Instruments and Methods"

Объ... ...ут
...
БИБЛИОТЕКА

Злоказов В.Б. P10 - 10855

Программа для построения оценок параметра экспоненциального распределения в условиях малой статистики

Приводятся постановка задачи построения оценок параметра экспоненциального распределения и их доверительных интервалов в условиях малой статистики, а также способы ее решения.

Программа, реализующая изложенные алгоритмы, была использована в экспериментах по синтезу элементов 104, 106, 107 (ОИЯИ, Дубна).

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Препринт Объединенного института ядерных исследований. Дубна 1977

Zlokazov V.B. P10 - 10855

A Program for the Estimation of the Parameter of an Exponential Distribution under the Conditions of Low Statistics

A problem formulation and algorithms to obtain estimates and their confidence intervals for the parameter of an exponential distribution under conditions of low statistics are given. The program, which implements the algorithms described, was applied to the analysis of data of the experiments on the synthesis of 104, 106, and 107 elements.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Preprint of the Joint Institute for Nuclear Research. Dubna 1977

© 1977 Объединенный институт ядерных исследований Дубна

Введение

Малая статистика регистрируемых продуктов распада некоторой активности делает невозможным эффективное использование техники регрессионного анализа для оценки периода полураспада этих активностей. Другие методы /а именно те, которые используют полную функцию распределения: метод моментов, метод максимального правдоподобия, байесовский метод и т.д.[1,2] дают более точные результаты; однако в общем случае существенное улучшение искомых оценок требует дополнительно сведения задачи к более простой форме.

Целью настоящей работы было выработать разумный компромисс между простотой формулировки задачи и ее реалистичностью, создать достаточно мощный аппарат для решения этой задачи, который был бы ориентирован на анализ данных с малой статистикой, и рассмотреть различные аспекты интерпретации получаемых оценок. ФОРТРАН-4 - программа, которая реализует описанный подход, была применена к анализу временных распределений треков в экспериментах по синтезу трансфермиевых элементов в Лаборатории ядерных реакций ОИЯИ[3].

1. Формулировка задачи для выбора малого объема

Пусть $\{t_j\}$ $j=1,\dots,M$ есть множество времен, в которых были зарегистрированы распады, и $P(x)$- функция распределения распада:

$$P(t<x) = 1 - \exp(-\ln 2x/T). \qquad /1/$$

Пусть (T_0,T_1) есть интервал регистрации. Условная вероятность того, что t_j меньше, чем t, при условии, что $t \in (T_0,T_1)$, равна:

$$P(t_j < t / t \in (T_0,T_1)) = P(t_j < t,\ и\ t \in (T_0,T_1)) / P(t \in (T_0,T_1)). \quad /2/$$

Вероятность того, что $t \in (T_0,T_1)$ и одновременно $t_j < t$, есть

$$P(t_j < t, t \in (T_0,T_1)) = \begin{cases} \exp(-\ln 2 T_0/T) - \exp(-\ln 2 t/T), & t \in (T_0,T_1), \\ 0, & t \notin (T_0,T_1). \end{cases} \quad /3/$$

Вероятность того, что $t \in (T_0,T_1)$, есть

$$P(t \in (T_0,T_1)) = \exp(-\ln 2 T_0/T) - \exp(-\ln 2 T_1/T). \quad /4/$$

Подставляя /3/ и /4/ в /2/, мы получим:

$$P(t_j < t / t \in (T_0,T_1)) = \frac{\exp(-\ln 2 T_0/T) - \exp(-\ln 2 t/T)}{\exp(-\ln 2 T_0/T) - \exp(-\ln 2 T_1/T)}. \quad /5/$$

Часто используется вращательная схема регистрации, когда времена распада регистрируются в цикле. Тогда вероятность того, что t_j меньше, чем t, равна:

$$P(t_j < t) = \sum_{n=0}^{\infty} (\exp(-\ln 2 (T_0 + T_1 n)/T) - \exp(-\ln 2 T_1 (n+1)/T)). \quad /6/$$

Используя формулу

$$(1 - x)^{-1} = \sum_{n=0}^{\infty} x^n \qquad \text{для} \quad |x| < 1,$$

мы можем показать, что /6/ сводится к /5/, так что далее мы не будем различать эти схемы и будем использовать только /5/.

Следующая физическая модель для измерения распада представляется достаточно реалистической. Имеется интересующий нас основной распад, которому мы припишем период полураспада T /в выражении /5//, и имеются 2 распада с существенно меньшим и большим полупериодами T_a и T_b, которые образуют фон. В частных случаях один из этих фонов /или оба/ могут отсутствовать. Тогда вероятность того, что зарегистрированное время /далее называемое треком/ t_j меньше, чем t, равна:

$$P(t_j < t) = q P_T(t_j < t) + q_a P_a(t_j < t) + q_b P_b(t_j < t), \quad /7/$$

где q, q_a, q_b являются вероятностями того, что j-й трек принадлежит распаду основного объекта, левого или правого фона соответственно, а P_T, P_a, P_b суть соответствующие функции распределения, каждая из которых относится к типу /5/ и зависит от параметров T, T_a, T_b соответственно.

Конечно, мы можем рассмотреть задачу оценивания всех величин: q, q_a, q_b, T, T_a, T_b, в предположении, что они неизвестны. Однако такая задача имеет мало шансов на успешное решение, если статистика имеющихся данных невелика. Наиболее вероятно, что такое решение будет давать сильно коррелированные оценки с огромными дисперсиями. Для того, чтобы упростить задачу, используем следующий подход:

1/ предполагается, что T_b много больше, чем T и T_a, и мы можем считать, что распределение P_b близко к равномерному распределению с постоянной плотностью $1/(T_1-T_0)$;

2/ предполагается /или требуется/, что/бы/ величины q, q_a, q_b /были/ известны /более или менее точно/ и фиксированы;

3/ затем предполагается, что либо T_a/T известно и нам надо оценить T, либо известно T_a и нам надо оценить отношение $R = T_a/T$; если $q_a = 0$, тогда нам надо оценить только T.

Такая формулировка проста, т.к. задача сводится к оцениванию одного параметра, и достаточно реалистична, поскольку возможности выполнить требования /1-3/ существуют часто.

2. Оценивание основного параметра

Таким образом, мы окончательно имеем следующую функцию распределения:

$$P(t_j < t) = qP(T) + q_a P(TR) + q_b t/T \quad , \qquad /8/$$

где функция распределения $P(X)$ имеет следующий вид:

$$P(X) = \frac{\exp(-T_0 \ln 2/X) - \exp(-t \ln 2/X)}{\exp(-T_0 \ln 2/X) - \exp(-T_1 \ln 2/X)} \; .$$

Используя /8/, мы должны оценить либо T, либо R.

1. Оценка метода моментов /M. M. E./

Поскольку

$$\int_{T_o}^{T_1} t dP(t_j < t) = qT/C(T) + q_a TR/C(TR) + q_b T_1/2 = f(T,R) \, ,$$

где $C(T) = \ln 2(\exp(-T_0 \ln 2/T) - \exp(-T_1 \ln 2/T))$, то вычисляя выборочное среднее множества треков $\{t_j\}$ /пусть это будет t_m/, мы имеем уравнение

$$t_m = f(T,R) \, .$$

Его решение даст оценку T /или R /. Аналогично /опуская подробные формулы, т.к. они слишком громоздки/ получаем:

$$\int_{T_o}^{T_1} t^2 dP(t_j < t) = g(T,R) \, .$$

Тогда уравнение для оценки дисперсии имеет следующий вид:

$$g(T,R) - f^2(T,R) = v \, ,$$

где v - выборочная дисперсия множества $\{t_j\}$. Этот метод позволяет получить оценки очень быстро.

2. Оценка максимального правдоподобия (M.L.E.)

Оценка дисперсии, получаемая использованием MME, часто недостаточно хороша; кроме того, если статистика мала, распределение оценок очень асимметрично и доверительный интервал, построенный с помощью среднего и дисперсии, не дает нужного представления о наиболее вероятном расположении оценок параметров. Поэтому MLE дает в общем случае более хорошие результаты, чем MME.

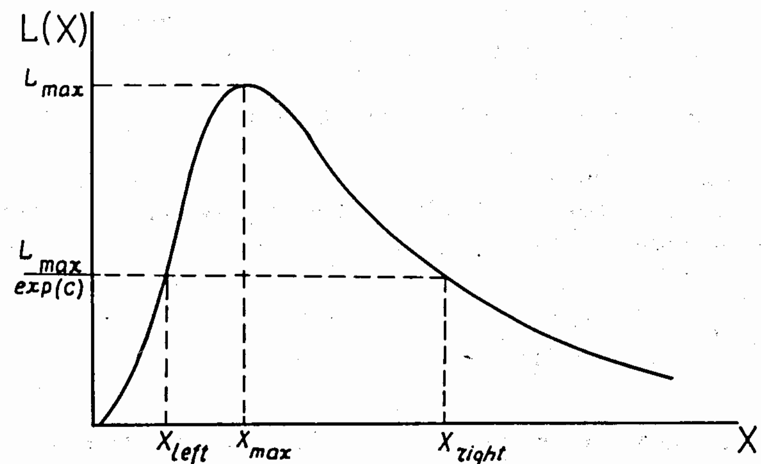
Функция правдоподобия записывается в следующем виде:

$$L(T,R) = \prod_{j=1}^{M} p(t_j) \, ,$$

где $p(t_j)$ /плотность вероятности в точке t_j / равна:

$$q(\ln 2)^2 \exp(-t_j \ln 2/T)/TC(T) \; +$$

$$+ \; q_a(\ln 2)^2 \exp(-t_j \ln 2/RT)/RTC(RT) + q_b/T_1 \, ,$$



Интервал равных шансов правдоподобия.

и мы ищем оценки T /или R/ из условия: максимизировать $L(T,R)$ на множестве значений T /или R /. При условии малой статистики распределение MLE будет достаточно асимметричным. Поэтому для построения доверительного интервала мы должны найти такие числа T_{lef}, T_{rig} /соответственно R_{lef}, R_{rig} /, что

$$\ln L(T_{lef}) = \ln L(T_{rig}) = \ln L(T_{max}) - C$$

или

$$\ln L(R_{lef}) = \ln L(T_{rig}) = \ln L(R_{max}) - C,$$

где T_{max}, R_{max} - оценки максимального правдоподобия T /или R / /интервалы равных шансов правдоподобия, см. /2/ /.

Константа C может быть вычислена исходя из условия, что для заданной вероятности p /уровень надежности/ выполняются равенства

$$\int_{T_{lef}}^{T_{rig}} L(T)dT = p, \quad \int_{R_{lef}}^{R_{rig}} L(R)dR = p, \qquad /9/$$

или вследствие вычислительной сложности /9/- путем моделирования с использованием генератора случайных чисел, распределенных по экспоненциальному закону типа /5/ с известными $T(R)$.

3. Байесовская оценка /BE/

Для исключительно низкой статистики /1-5 треков/ можно построить BE. В нашем случае она получается из условия

$$T_B = \frac{\int TL(T)dT}{\int L(T)dT}, \quad R_B = \frac{\int RL(R)dR}{\int L(R)dR}. \qquad /10/$$

Области интегрирования в /10/ могут быть сведены к интервалам, которые почти наверное содержат искомое

значение /такое сведение делает оценку более устойчивой к грубым ошибкам в данных/.

В этом случае можно также в качестве доверительного интервала построить интервал равных шансов правдоподобия. Байесовская оценка в общем случае смещена, однако величина смещения невелика и стремится, как правило, к нулю, когда объем выборки возрастает неограниченно.

В простом случае /один источник распада, нет фона, $T_j \gg T$/ оценки имеют аналитическое выражение:

$$MME = MLE = \sum_{j=1}^{n} t_j \ln 2/n; \quad BE = \sum_{j=1}^{n} t_j \ln 2/(n+1).$$

3. Описание программы

Для реализации вышеописанного подхода была написана программа HTIME на языке **ФОРТРАН-4**. Входными данными являются:

1/ Q, Q_1, Q - число треков в каждой компоненте;
T_1 - верхняя граница интервала измерения;
C - константа для построения интервала равных шансов правдоподобия /для уровня надежности 67% $C \approx 1$; для уровня надежности 90% $C \approx 1{,}15$/;

2/ либо значение T_a, и тогда оценивается R; либо значение R, и тогда оценивается T;

3/ треки /в реальной временной шкале, T_0 вычтено/.

Программа строит MME, которая рассматривается как начальное значение для требуемого периода полураспада /или отношения/, затем, используя это начальное значение, строится MLE. Дополнительно HTIME может включать в себя подпрограмму для построения BE. Вычисляются доверительные интервалы /также интервалы равных шансов правдоподобия/, значения критерия χ^2 и критерия Колмогорова. Эти значения используются, чтобы проверить, не противоречат ли треки гипотезам о распределениях /8/ с T /или R /, равными полученным оценкам.

Примеры применения HTIME к анализу данных в экспериментах по синтезу элементов 104, 106 и 107 даны в работах /3,5,6/.

Автор глубоко благодарен профессору Ю.Ц.Оганесяну за постановку задачи. Он хотел бы также поблагодарить д-ра А.Г.Демина /ОИЯИ/ и д-ра Й.Ничке /Лаборатория им. Э.Лоуренса, Беркли/ за полезные обсуждения многих физических аспектов задачи.

Литература

1. *Яноши Л. Теория и практика обработки результатов измерений. М., Мир, 1968.*
2. *Дженкинс Г.М., Ваттс Д.Г. Спектральный анализ и его приложения. М., Мир, 1971.*
3. *Друин В.А. и др. ОИЯИ, Е7-10499, Дубна, 1977.*
4. *Wilks S.S. Mathematical Statistics, John Wiley & Sons, Inc., New-York, London, 1962.*
5. *Друин В.А. ОИЯИ, Р7-10359, Дубна, 1977.*
6. *Оганесян Ю.Ц. и др. ОИЯИ, Д7-9866, Дубна, 1976.*

Рукопись поступила в издательский отдел
14 июля 1977 года.