

ВЫЧИСЛЕНИЕ ПЕРИОДА ПЕРЕМЕННОЙ ЗВЕЗДЫ



Иван Леонидович Андронов

Что такое периодический процесс?

Человеческому разуму свойственен поиск закономерностей в окружающем мире, что позволяет более или менее точно прогнозировать будущее. К тому же многие процессы во Вселенной имеют периодическую природу. Например, Земля вращается вокруг оси с почти постоянным периодом, и периодически изменяется высота Солнца над горизонтом, хотя яркость света зависит не только от времени года, но и сложного характера изменения погоды. Но, несмотря на одновременность влияния различных механизмов, выявление скрытых периодичностей является одним из первых типов анализа любого сигнала. С точки зрения математики, периодическим называется процесс, для которого можно записать уравнение:

$$x(t + P \cdot E) = x(t), \quad (1)$$

где

$x(t)$ - значение сигнала x в момент времени t ; P называется периодом; целое значение E - номером цикла.

Конечно, периодом этого процесса может быть не только P , но и любое кратное значение $j \cdot P$, где j - целое. Поэтому договариваются периодом называть наименьшее значение P , для которого выполняется соотношение (1).

Пример такого процесса показано на рис. 1. Период легко найти как наименьший промежуток времени между повторяющимися элементами кривой - минимумами, максимумами, пересечениями восходящей или нисходящей ветвью кривой фиксированного уровня - среднего или любого другого значения.

Но непрерывное значение времени t в уравнении (1) является идеализацией, ибо в реальности мы имеем отдельные наблюдения только в некоторые моменты, между которыми наблюдений не было - или вследствие плохой погоды, солнечного света, или нехватки времени наших наблюдений на телескопе с перенапряженной научной программой. Не говоря уже о точности наблюдений, которая приводит к тому, что измерения даже постоянного сигнала, в общем случае, различаются. Так что точно уравнение (1) не выполняется даже при наиболее благоприятных реальных условиях. Поэтому для определения периода необходима разработка математических методов, учитывающих как ограниченность ряда наблюдений, так и погрешности измерений сигнала.

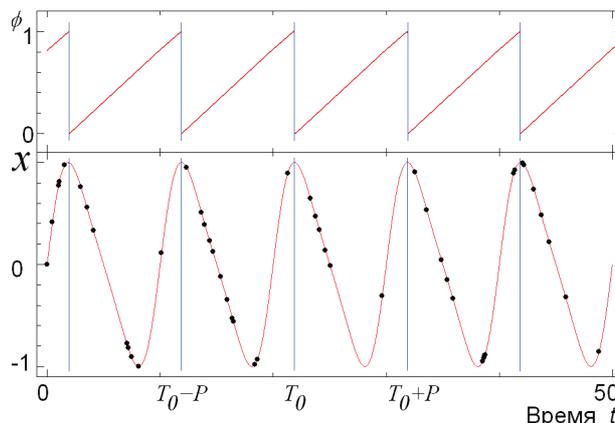


Рис. 1. Сверху: изменение фазы со временем. Снизу: «кривая блеска» с периодом $P = 10$, сгенерированный на компьютере для случайных моментов времени. Если не смотреть на «теоретическую» кривую блеска (линия), то точки наблюдений «кажутся совсем случайными». Но, при приведении к фазовой кривой, показывают хорошую переменность.

На верхней части рисунка показана зависимость фазы ϕ_1 от времени линейный рост от нуля до единицы, затем мгновенное падение вновь до нуля, и так далее.

От бесконечного времени до периодической фазы

Самым естественным является решение, согласно которому время любого наблюдаемого сигнала нужно сместить на соответствующее целое число периодов

$P \cdot E$ так, чтобы расстояние между первым и последним таким «наблюдением» во времени не превышала период P . Такое значение $\tau_j = (t_j - (T_0 + P \cdot E_j))$ для наблюдения

номер j называется «возрастом». Здесь T_0 называется «начальной эпохой», которая соответствует возрасту, равному нулю. В повседневной жизни «возраст» - это время, прошедшее после полуночи, которое изменяется от 0 до «почти» 24 часов. «Почти», потому что через секунду после $23^{\text{h}}59^{\text{m}}59^{\text{s}}$ (23 часов 59 минут 59 секунд) время станет равным не $24^{\text{h}}00^{\text{m}}00^{\text{s}}$, а $0^{\text{h}}00^{\text{m}}00^{\text{s}}$, но при этом изменится календарная дата (номер дня).

Но периоды изменения сигнала бывают очень разнообразны, поэтому для кривой зависимости $x_j(\tau_j)$ которую в звездной фотометрии называют «кривой блеска», нужно каждый раз выбирать собственный масштаб. Чтобы унифицировать эту процедуру, договорились измерять «возраст» в единицах периода, т.е. ввести так называемую «фазу» $\phi_j = \tau_j / P = \tau_j f$, где величина $f = 1/P$, обратная периоду P , называется частотой. В физике принято измерять время в секундах, а частоту в герцах ($=1/c$). В астрономии время традиционно измеряется в сутках, а частота в «циклах в сутки».

Но и фаза может быть неоднозначно определенной величиной, так как разные авторы могут отсчитывать время от различных явлений - кто-то от начала зры, кто-то от полудня в Гринвиче, кто от начала собственных наблюдений.

Поэтому договорились ввести еще одну характеристику - уже упоминавшуюся «начальную эпоху» T_0 . Окончательно фаза определяется, как $\phi_j = v_j - E_j$, где $v_j = (t_j - T_0) / P$, $E_j = \text{int}(v_j)$, где $\text{int}(v)$ - наибольшее целое, не превышает значения v . Например,

$\text{int}(-4.7) = -5$, $\text{int}(+4.7) = 4$. При таком определении, $0 < \phi_l < 1$, а зависимость $x_l(\phi_l)$ называется «фазовой кривой» (блеска или иного типа сигнала). Величина ν не имеет специального названия, но она имеет важное значение в теории поиска периода, поэтому назовем ее «полной фазой» (что является суммой классической фазы ϕ и номера цикла E_j).

Зависимость фазы от времени показана на рис. 2. Она имеет характерный «пилообразный» вид с разрывом в целых значениях аргумента. Иногда интервал от нуля до единицы называют «основным интервалом определения фазы», потому что, естественно, можно фазовую кривую повторять бесконечное количество раз вправо и влево. Зависимость $x_j(\nu_j)$ имеет период, равный единице. Значение ν , например, -0.1 , 0.9 , -121312312.1 и 290663.9 соответствуют одной и той же фазе $\phi_j = 0.9$.

С «возрастом» мы имеем дело каждый день, потому что время на наших часах - это отрезок времени от полуночи, который увеличивается до 24 часов, а затем становится равным нулю полуночи. А вот если бы мы разделили этот возраст на период в 24 часа, то получили бы фазу. Вместо предложения «начало рабочего дня в 9-часов» писали бы «начало рабочего дня в фазе $0.375 (= 9/24)$ ». Такой же самый результат, только непривычно в быту. И естественно в астрономии.

Что такое «хорошая» фазовая кривая?

Если период использовано правильный (обозначим его P_0), то фазовая кривая блеска будет показывать какую-то линию с рассеянием согласно точности наблюдений (рис. 2). Для пробных периодов, кратных настоящему $P = jP_0$, кривая показывает j повторений вместо одного. Если период выбран такой, который немного отличается от настоящего, то рассеяние увеличится за счет «рассеяния» наблюдений по фазам, но кривая еще будет заметна. Когда это отклонение достигнет критического значения, то фазовая кривая будет напоминать случайное распределение точек в прямоугольнике.

В 1960-е годы даже была идея показывать фазовую кривую блеска на осциллографе, и, подкручивая колесико регулировки, подбирать необходимую частоту, чтобы кривая была лучшей на вид. Аналогично мы настраивали старые радиоприемники, подбирая частоту, при которой сигнал «наилучший». В современных электронных устройствах поиск частот радиостанций и настройка проводятся автоматически.

Принципиально, идея «осциллографа» неплохая, но для очень ограниченного класса наблюдений, когда их интервал лишь в несколько раз превышает период, или когда мы должны менять частоту не очень много раз. Но, предположим, что мы исследуем звезду типа RR Лир с периодом около 0.5 суток. За год будет ~730 полных периодов колебаний (астрономы называют их «циклами»). За 40 лет наблюдений в «звездном патруле» в обсерваториях в Зоннеберге (Германия) или Одессе пройдет около 30 тысяч циклов. Но сколько точно? 30000, 30001, 29663 или иное число, к тому же, возможно, не целое?

Поэтому наибольшую популярность приобрела идея автоматического поиска периодов, которые соответствуют лучшим фазовым кривым, с помощью компьютера. Понятно, что компьютеру нужны конкретные математические правила, по которым определяется «качество» фазовой кривой, а не общие слова об эстетической красоте или соответствии формы уже известным типам переменных звезд.

Таким критерием может быть какая-то функция θ от частоты f , начальной эпохи T_0 и набора наблюдений $(x_\nu t_\nu, \nu = 1 \dots N)$: $\theta(f, T_0; x_\nu t_\nu)$. Все методы поиска периода отличаются только конкретным видом такой зависимости, называемой «тест-функцией», или (редко) «пробной функцией», «целевой функцией».

В этой зависимости наблюдения являются постоянными (константами), не зависят от «пробной частоты» f или начальной эпохи T_0 . Вообще, тест-функция не должна зави-

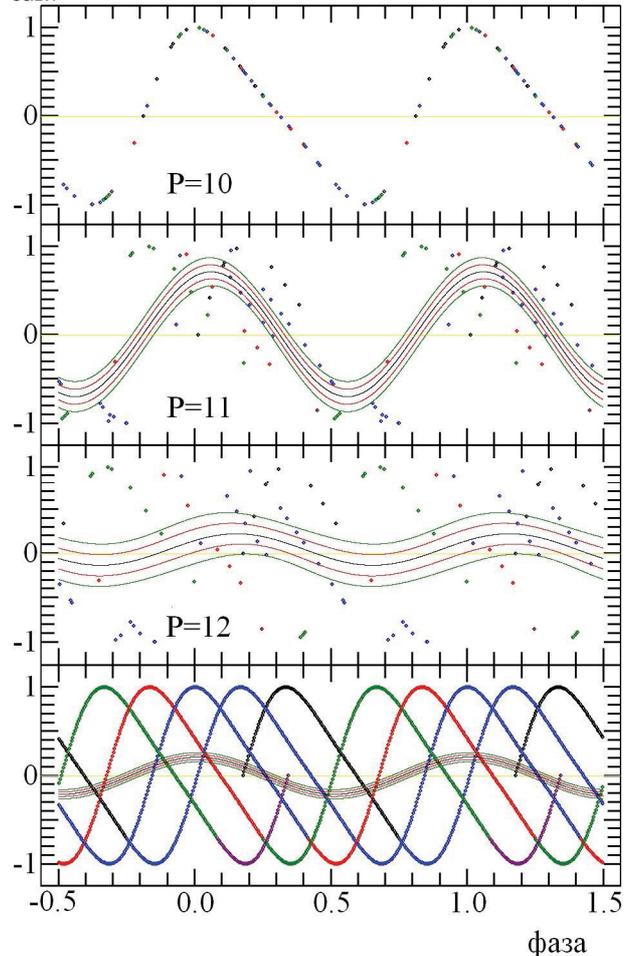


Рис. 2. Фазовые кривые для сгенерированных наблюдений для четырех значений пробного периода: $P=10$ для «настоящего» периода мы видим, что «наблюдения» находятся на кривой, «вытянутой в ниточку»; для удвоенного периода (не показано) наблюдается двойная волна переменности, иногда она и есть настоящая, например, в затменно-двойных системах типа W Большой Медведицы; $P = 11$ - период близок к настоящему, плавная переменность заметна, но уже присутствует сдвиг фаз, что приводит к рассеиванию кривой блеска (для сравнения программой MCV проведена сглаживающая синусоида и «коридоры ошибок» на уровне $\pm 1\sigma$ и $\pm 2\sigma$); $P = 12$ сдвиг фаз настолько велик, что уже невозможно догадаться, что сигнал действительно периодический, так как выглядит, как случайный (кривая блеска показывает сильное рассеяние, однако, формально сглаживающую функцию можно рассчитать, а ее погрешность σ очень велика); на нижнем рисунке для пробного периода $P = 12$ показаны не сгенерированные «данные», а 1000 равномерно расположенных точек по всему интервалу (когда наблюдений много, то на кривой блеска наблюдаются последовательности, отмеченные разными цветами).

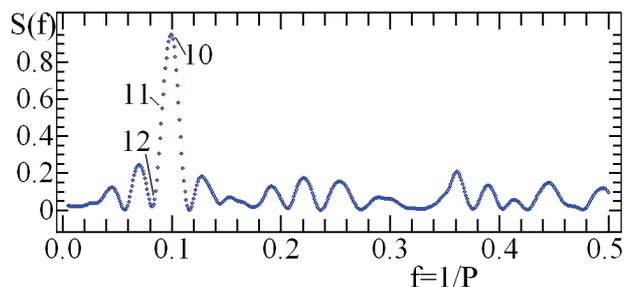


Рис. 3. Периодограмма $S(f)$ как зависимость тест-функции $S(f)$ от частоты $f=1/P$ для синусоидальной аппроксимации. Самый высокий максимум наблюдается вблизи «правильного» значения $P=10$, однако, чуть-чуть отличается

от него, поскольку кривая блеска периодическая, но несинусоидальная (мульти-гармоническая). Цифрами показаны значения периодов, использованные для построения фазовых кривых на рис. 2. Наблюдается много мелких пиков, их ширина примерно одинакова, а средняя высота "шумов" $\langle S \rangle \sim 2/(n-1)$, где n - число наблюдений.

сеть от начальной эпохи T_0 , поскольку выбрать эту величину мы можем по своему усмотрению, даже не связывая ее с физическими процессами в звезде (например, от полудня по гринвичскому, киевскому или московскому времени).

То есть лучшей будет функция $\theta(f) = \theta(f, T_0; x_v, t_v)$, зависящая только от одной переменной – частоты $f (= 1/P)$. Зависимость $\theta(f) = \theta(1/P)$ тест-функции θ от пробной частоты f (или периода P) называется "периодограммой".

Как найти компромисс?

Рассмотрим случай, когда фазовая кривая рассматривается для двух значений периода P_1 и P_2 . Фаза для одного и того же наблюдения изменится на величину $v_j(P_2) - v_j(P_1) = (t_j - T_0)/P_2 - (t_j - T_0)/P_1 = (t_j - T_0)(1/P_2 - 1/P_1) = (t_j - T_0)(f_2 - f_1)$.

То есть при изменении периода (или соответствующей частоты), фаза каждого наблюдения изменяется пропорционально расстоянию от начальной эпохи T_0 .

Можно сравнить с ситуацией, когда расстояние автомобиля от точки одновременного старта меняется пропорционально скорости. Или с разбеганием галактик согласно закону Хаббла. В этом сравнении есть и свои последствия: нет центра, то есть каждую точку можно считать центром, от которого проходит разбегания со скоростью, пропорциональной расстоянию от центра.

Относительно фазовых кривых, то можно выбрать из бесконечного количества возможностей начальную эпоху так, чтобы она равнялась времени какого-либо из наблюдений, например, первого t_1 . Тогда полная фаза последнего наблюдения $v_N = (t_N - t_1) \cdot f$ будет расти пропорционально частоте f . Поэтому при периодограммном анализе вычисления делают с постоянным шагом Δf по частоте f , а не по периоду P . Как правильно выбрать этот шаг? Чем меньше он будет, тем меньше будет сдвиг фаз между первым и последним наблюдениям, и тем подробнее можно наблюдать изменения фазовой кривой блеска, и тем точнее выбрать "лучшую" кривую и соответствующие ей значения частоты и периода. Но тем большее число таких пробных частот, т.е. времени расчетов. Но, сокращая объем вычислений, можно пропустить лучшую фазовую кривую, если ее частота окажется внутри интервала между двумя пробными частотами. То есть нужен компромисс между детальностью периодограммы и объемом вычислений. Можно порекомендовать выбрать шаг $\Delta f = \Delta\phi / (t_N - t_1)$, где величина изменения разности фаз $\Delta\phi$ между последним и первым наблюдениям при изменении частоты на величину Δf не превышает некоторого критического значения $\Delta\phi_{\max}$. Эта величина $\Delta\phi_{\max}$ различна для разных методов, но вообще выбирается от $1/N$ (а это маленькая величина для большого объема наблюдений) до $\sim 0,05-0,10$. Напр., в программе MCV при введении числа пробных частот высвечивается значение $\Delta\phi_{\max}$ и вычисление периодограммы запрещено, если $\Delta\phi_{\max} > 0,1$, чтобы избежать «пропуска» «лучшего» экстремума.

"Номенклатура" методов

Все методы периодограммного анализа можно разделить на две большие группы - так называемые "параметрические" и "непараметрические" методы. В первой группе фазовая кривая блеска приближается любой

функцией с периодом, который равен единице, например, синусоидой

$$x_c(\phi) = C_1 + C_2 \cdot \cos(2\pi\phi) + C_3 \cdot \sin(2\pi\phi).$$

Затем для каждого значения пробной частоты f методом наименьших квадратов определяются коэффициенты (параметры) C_1, C_2, C_3 , а критерием согласованности фазовой кривой с этой аппроксимацией является сумма квадратов "невязок" (или отклонений наблюдений от сглаживающей функции, рассчитанной для этого же аргумента) $(O - C)_k = x_k - x_C(\phi_k)$. В этом примере,

$$x_k - (C_1 + C_2 \cdot \cos(2\pi\phi_k) + C_3 \cdot \sin(2\pi\phi_k)):$$

$$\Phi(f) = \sum_{k=1}^N (x_k - x_C(\phi_k))^2 = (x_1 - x_C(\phi_1))^2 + (x_2 - x_C(\phi_2))^2 + \dots + (x_N - x_C(\phi_N))^2$$

Чем меньше $\Phi(f)$, тем, в среднем, наблюдения ближе к аппроксимации, и лучшая в этом смысле частота соответствует наименьшему значению $\Phi(f)$.

В "параметрическом" периодограммном анализе часто вместо минимума этой функции ищут максимум функции $S(f) = 1 - \Phi(f)/\Phi_0$, где Φ_0 является суммой квадратов отклонений наблюдений от их среднего значения $\langle x \rangle = (x_1 + x_2 + \dots + x_N)/N$

$$\Phi_0 = \sum_{k=1}^N (x_k - \langle x \rangle)^2 = (x_1 - \langle x \rangle)^2 + (x_2 - \langle x \rangle)^2 + \dots + (x_N - \langle x \rangle)^2$$

Именно тест-функция $S(f)$ и приведена на рис. 3.

Величина $\Phi_0 = N\sigma_0^2$, где σ в математической статистике называется "смещенной оценкой среднеквадратического отклонения наблюдений от среднего значения", или, короче, среднеквадратичным отклонением. "Параметрические" методы иногда называют методами группы "точка-кривая", потому тест-функция зависит от среднеквадратического расстояния наблюдаемых точек от аппроксимирующей (дословно с латыни, "приближающей", в плохих компьютерных переводчиках, "приспосабливающейся", но по-русски, «сглаживающей») кривой. Для периодических процессов несинусоидальную кривую можно аппроксимировать тригонометрическим полиномом степени s (иногда называемым "усеченной суммой ряда Фурье")

$$x_C(\phi) = C_1 + C_2 \cdot \cos(2\pi\phi) + C_3 \cdot \sin(2\pi\phi) + C_4 \cdot \cos(4\pi\phi) + C_5 \cdot \sin(4\pi\phi) + \dots + C_{2s} \cdot \cos(2s\pi\phi) + C_{2s+1} \cdot \sin(2s\pi\phi).$$

Теоретически количество членов суммы "ряда Фурье для периодической функции" должно быть бесконечно, но практически ограничиваются не очень большой величиной s , поскольку дальнейшие члены суммы очень малы по сравнению с точностью наблюдений и не являются статистически значимыми.

Периодическая волна

$$"+C_{2s} \cdot \cos(2j\pi\phi_1) + C_{2s+1} \cdot \sin(2j\pi\phi_1)"$$

называется $(j-1)$ гармоникой основной частоты, поэтому сумму называют "мультигармоническим приближением". А тригонометрический полином степени, например, 4 можно назвать "приближением с 3 (= 4-1) гармониками". Следует отметить, что изменение аргумента синуса или косинуса целое число раз на 2π не меняет значение соответствующей функции. Аргумент

$$2\pi v_j = 2\pi(t_N - T_0) \cdot f_j = 2\pi(t_N - T_0) \cdot (jf) = 2j\pi v_1$$

соответствует фазе для частоты $f_j = jf_1$, т.е. периоду $P_j = 1/(jf) = P/j$.

Такое соотношение между периодами $P_j = P/j$ не является уникальным. Например, когда мы определяем уравнение времени, то есть разницы между прямыми прохождениями "настоящего" и "среднего" Солнца, то мультигармонические приближения являются

реалистичными, потому что период в один год соответствует эллиптичности орбиты Земли (а его гармоники повышают точность такого приближения), а проекция эклиптики на небесный экватор дает полугодовую периодичность. Но если мы возьмем, например, видимые с поверхности Земли координаты Луны, то нужно учитывать по меньшей мере три периода: вращения системы "Земля-Луна" вокруг Солнца P_1 , сидерический (относительно звезд) орбитальный период Луны P_2 и синодический (относительно Солнца) период вращения Земли вокруг своей оси (что приводит к суточному параллаксу, т.е. смещения видимого положения небесного тела из-за движения наблюдателя) P_3 . Отношение этих периодов не являются целыми числами ($P_1/P_3 = 365.2422$, $P_2/P_3 = 27.32$, $P_1/P_2 = 13.369$), поэтому мульти-гармонические приближения не являются оптимальными. Процессы, в которых присутствуют колебания с несколькими периодами (частотами $f_1 \dots f_s$), называют мультипериодическими. Их математическую модель можно записать как

$$x_C(t) = C_1 + C_2 \cdot \cos(2\pi(t - T_0) \cdot f_1) + C_3 \cdot \sin(2\pi(t - T_0) \cdot f_1) + \dots + C_{2s} \cdot \cos(2\pi(t - T_0) \cdot f_s) + C_{2s+1} \cdot \sin(2\pi(t - T_0) \cdot f_s)$$

Хотя вообще на отношение f_j/f_k нет ограничений, часто некоторые из этих величин могут быть целыми, а некоторые периодические составляющие могут быть мультигармоническими. Бывают и случаи, когда разница между частотами является постоянной величиной (например, в астросейсмологии).

Кроме синусоид, для анализа могут быть использованы другие периодические функции, например, зависимость между фазой и координатой или скоростью в случае эллиптической орбиты (визуально, затменно- или спектрально-двойные звезды), или функции с лучшими аппроксимирующими свойствами - так называемые сплайны.

Специальный вид аппроксимации кривых блеска затменных звезд типа Алголя, характеризующихся узкими минимумами, который в ~2 раза точнее тригонометрических полиномов, недавно рассмотрен нами (И.Л.Андронов, 2012) Но это уже тема для следующих статей.

"Непараметрический" метод Лафлера-Кинмана-Холопова

Когда мы аппроксимируем, например, синусоидой, кривую блеска с несинусоидальной кривой, то можем получить ошибочные результаты. Например, кривая блеска затменной звезды типа Алголь имеет достаточно узкий минимум - затмение, которое на синусоиду совсем не похоже. Можно аппроксимировать тригонометрическим полиномом с большой степенью, но это формальное приближение может привести к появлению ложных пиков на периодограмме. И на кривой появляются "волны" с малыми периодами P/j . Не говоря уже, что мы "навязываем Природе наше представление о ней".

Поэтому были разработаны альтернативные "непараметрические" методы, в которых "качество" фазовой кривой зависит от расстояния между ближайшими по фазе точками. Поэтому такие методы относят к так называемой группе "точка-точка". Первый из таких методов был предложен Лафлером и Кинманом в статье 1965 года. Его основная идея заключалась в том, что для рассеяния фазовой кривой был избран квадрат разности между последовательными по фазе наблюдениями:

$$\Phi(f) = (y_2 - y_1)^2 + (y_3 - y_2)^2 + \dots + (y_N - y_{N-1})^2.$$

Если кривая имеет малое рассеяние (как говорят, "в ниточку"), то эта величина будет небольшой по сравнению с рассеянием для неправильного периода. Отметим, что для каждого значения пробной частоты за моментами наблюдений t_k , определяются фазы ϕ_k , вводится «рабочий» массив $y_k = x_k$, затем пары точек (y_k, ϕ_k) сортируются в соответствии с ростом фаз ϕ_k . То есть каждой точке (x_i, t_i) на фазовой кривой соответствует одна точка

(y_k, ϕ_k) , но ее номер k различен для каждого значения пробной частоты, поскольку сортировка проводится по фазам, а не по времени.

Но к сожалению, в этой тест-функции не выполнено условие "равенства перед законом" всех точек наблюдения, потому значения x_1 и x_N входят только в один член суммы, в то время как другие - в два. Кроме того, «кто первый, а кто последний» зависит как от пробной частоты f , так и от начальной эпохи T_0 .

Это неравноправие было исправлено московским астрофизиком П. Н. Холоповым, который в 1969 году опубликовал алгоритм усовершенствованного метода и соответствующую компьютерную программу. С тех пор к тест-функции Лафлера и Кинмана был добавлен член $(x_1 - x_N)^2$:

$$\Phi(f) = (x_2 - x_1)^2 + (x_3 - x_2)^2 + \dots + (x_N - x_{N-1})^2 + (x_1 - x_N)^2.$$

Действительно, поскольку фазовая кривая имеет период, равный единице, то мы можем продолжить фазовую кривую, формально написав $\phi_{N+1} = 1 + \phi_1$, $x_{N+1} = x_1$,

поэтому следующей по отношению к фазе ϕ_N будет точка с фазой $(1 + \phi_1)$.

Чтобы было легче сравнить тест-функцию с ожидаемым для полностью аperiodического сигнала значением $2\Phi_0$, тест-функцию можно записать в безразмерном виде:

$$\Theta(f) = \Phi(f) / \Phi_0$$

Так что на периодограмме будет рассеяние у "aperiodической" величины $\Theta(f) = 2$, а у "настоящего" значения частоты f_0 будет наблюдаться минимум. Чем глубже такой минимум, тем меньше среднее отклонение последовательных точек друг от друга, тем уже полоса кривой блеска, тем лучше «качество кривой». И тем ниже вероятность "ложной тревоги", то есть того, что такой глубокий минимум на периодограмме будет наблюдаться вследствие случайно малого рассеивания кривой блеска на какой-либо из большого числа пробных частот, когда на самом деле в сигнале этой частоты нет. Но случай и есть случай, поэтому возможны "красивые" фазовые кривые для "ненастоящего" периода для данного набора данных, которые потом не подтверждаются последующими наблюдениями.

Практическое применение метода

Алгоритм Лафлер-Кинмана в модификации П. Н. Холопова очень активно применяется для поиска периода переменных звезд и много сотен, или даже тысяч значений, приведенных в "Общем Каталоге Переменных Звезд" (Интернет-адрес <http://www.sai.msu.su/cgi-bin/wdbr95/gcvs/stars/form>), получены с помощью его программы. Конечно, с появлением новых компьютеров этот алгоритм многократно переводился на Фортран, Паскаль, Бейсик, Си, Маткад и другие компьютерные языки, добавляя авторское понимание интерфейса и удобства работы с программой.

Описание одной из компьютерных реализаций метода, сделанной Виталием Бреусом, тогда ещё школьником - активным членом Одесского отделения Малой Академии Наук, ныне аспирантом. Поскольку эта программа предлагается другим любителям и опубликована на странице Украинской Ассоциации Наблюдателей Переменных Звезд (<http://uavso.pochta.ru>, <http://uavso.org.ua>), используем ее для расчетов примеров результатов применения метода с редактированием рисунков для журнальной публикации. На рис. 4-6 показаны различные виды периодограмм для модельных наблюдений, показанных на рис. 1. Как и следовало ожидать, при "неправильном" значении частоты значение тест-функции около 2, и уменьшается около «правильной» частоты $f_0 = 0.1$ и частот $f_j = f_0 / j$ (т.е. "кратных периодах" $P_j = jP_0$). Это отличает «непараметрические» методы от «параметрических» мультигармонических, где

кратными являются частоты, а не периоды. Но этого тоже следовало ожидать - период $2P_0$ или jP_0 также является периодом, только в нем фазовая кривая состоит из 2 или j почти одинаковых частей. Отметим, что в сумму входят разности значений последовательных точек на фазовой кривой блеска, а в эту разность входят не только статистические погрешности наблюдений, но и систематические разности из-за переменности сигнала. Напр., на рис.1 показан сигнал с нулевой статистической погрешностью, но его значение периодически возрастает и убывает, поэтому значение убывает не до нуля, а до некоторого малого значения $\Theta(f_1)$. При удвоенном периоде, средняя разность фаз последовательных точек удваивается, и поэтому значение $\Theta(f_2) \approx 2^2 \Theta(f_1) = 4\Theta(f_1)$. Для других кратных периодов, $\Theta(f_j) \approx j^2 \Theta(f_1)$, однако, такое возрастание замедляется по сравнению с этой формулой и потом останавливается, когда точек на одну волну (интервал фаз шириною $1/j$ становится мал). Напр. для использованного нами иллюстративного сигнала продолжительностью 50, использование пробных периодов более 50 вообще не приводит к изменению тест-функции $\Theta(f)$, поскольку остаток фаз просто не заполнен точками.

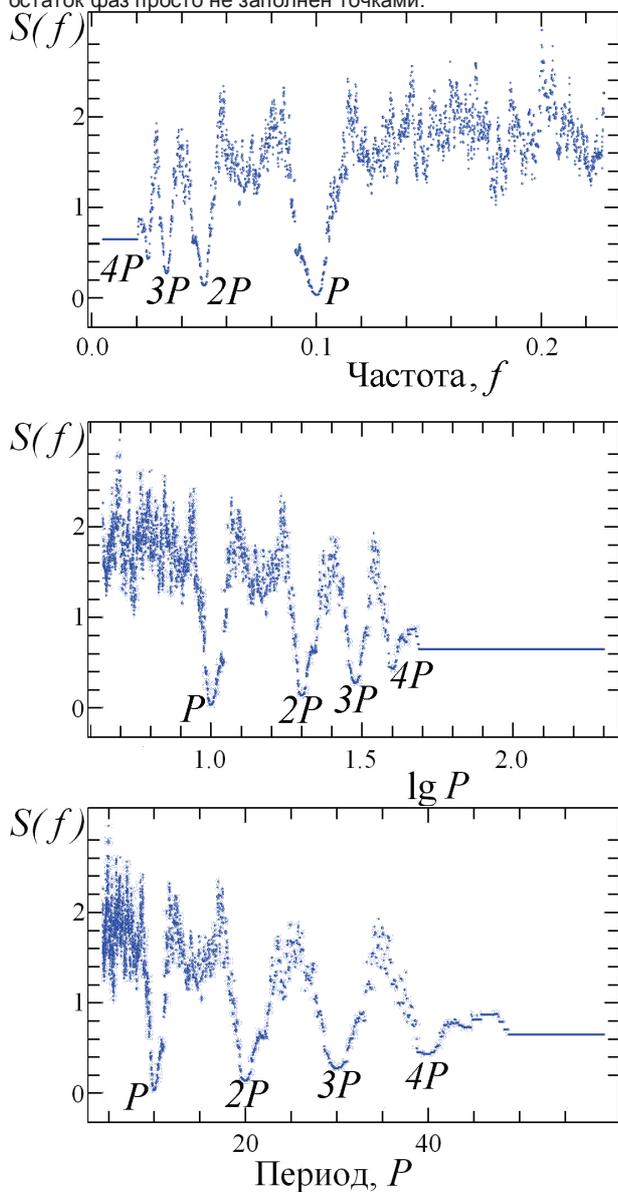


Рис. 4. Периодограмма для метода Лафлера-Кинмана-Холопова для различных значений аргумента – частоты $f=1/P$, логарифма периода $\lg P$ и периода P . В этом методе ширина пиков примерно постоянна для $\lg P$ (как и в случае всплеск-преобразования (wavelet transform), заслуживающего отдельной публикации). Отметим, что, при использовании синусоидальной аппроксимации и других

модификаций «преобразования Фурье» примерное «постоянство ширины» наблюдается для частоты f . Во всех методах, ширина пика растет с увеличением периода.

Разрывность периодограммы, или эффект перемешивания

Если внимательно присмотреться к периодограмме, то можно заметить, что она является разрывной. Т.е. небольшая разница в частоте может приводить к прыжку в значении. Это связано с тем, что точки даже с близкими фазами при изменении частоты могут изменить свои порядковые номера k , поэтому в сумму будут входить разности между значениями сигнала в разных парах. Наоборот, если номера не изменились, не изменится и значение тест-функции. Т.е. периодограмма будет «кусочно-постоянной».

Поскольку средняя разность фаз между последовательными наблюдениями равна $1/N$ (период фазы 1, разделенный на количество точек N), то это значение и можно предложить как $\Delta\phi$ в формуле для определения шага по частоте. Но в целом ширина минимума значительно больше $\Delta f_{\min} = 1/(t_N - t_1)$, что соответствует $\Delta\phi = 1$.

Следует заметить, что для «кратных» периодов $P_j = jP_0$, ширина пиков пропорциональна частоте (если по оси абсцисс будет частота), или периоду (если по оси абсцисс будет период), т.е. постоянной будет относительная ширина минимума на периодограмме.

С увеличением номера j глубина минимума падает медленно, пока число наблюдений N/j на одну «волну» на периодограмме еще достаточно велико. Но дальше в разницу сигнала будет существенно добавляться изменение за счет систематической фазовой изменчивости и глубина уменьшится до исчезновения минимума в «шуме» периодограммы.

Другие модификации метода

После опубликования первого алгоритма было предложено несколько методов модификаций, которые могли бы дать большую эффективность для более специализированных типов сигнала.

Все они базируются на понятии «расстояния» $U_k = U(\Delta x_k, \Delta\phi_k)$ между двумя точками на фазовой кривой, где $x_k = x_{k+1} - x_k$, $\Delta\phi_k = \phi_k + 1 - \phi_k$, и снова формально $\phi_{n+1} = 1 + \phi_1$, $x_{n+1} = x_1$. Периодограмма может быть записана, как $\Theta(f) = U_1 + U_2 + \dots + U_N$, вот только функции $U(\Delta x_k, \Delta\phi_k)$ будут разными. Перечислим их:

$U(\Delta x_k, \Delta\phi_k) = (\Delta x_k)^2$ (Lafler-Kinman (1965), Холопов, 1970), $|\Delta x_k|$ (Deeming, 1970), $|\Delta x_k|^\gamma$ (Пельт, 1980). В этой группе модификаций «расстояние» зависит только от сигнала, а не от фазы. При $\gamma = 1$ уменьшается относительный вклад наблюдений с большими отклонениями, которые могут даже привести к появлению минимума на периодограмме для совсем неправильного значения частоты. При больших значениях γ этот вклад будет увеличен, и Пельт не рекомендует давать значение, больше $\gamma = 4$, так как это приводит к увеличению шума на периодограмме.

Если две последовательные точки соединить отрезком, то его можно считать гипотенузой треугольника, два катета которого параллельны координатным осям. Вместо квадрата одного из катетов можно использовать квадрат гипотенузы, т.е. квадрат расстояния между точками на плоскости.

Тогда $U(\Delta x_k, \Delta\phi_k) = (\Delta x_k)^2 + (\Delta\phi_k)^2$, как предложил швейцарский ученый Burk с коллегами в 1970 для циклических процессов. Действительно, если обе величины Δx_k , и $\Delta\phi_k$ были бы нарисованы в одном масштабе, то

квадрат гипотенузы имел бы физический смысл. Но на самом деле фаза является безразмерной величиной (отношение промежутка времени к периоду), в то время как Δx_k может быть звездной величиной, лучевой скоростью, температурой, давлением или даже количеством денег в кошельке или расстоянием до дома. Поэтому результат зависит от масштаба измерения сигнала, т.е. будет разный, если мы, например, изменим единицы измерения с километров на метры. Поэтому следующие модификации уже учитывали этот факт и единица измерения сигнала выбиралась равной $2(x_{\max} - x_{\min})$, т.е. двойной амплитуде изменений сигнала (Dworetzky и другие 1980). Впрочем, Дворецкий предложил метод "наиболее короткой веревки", соединяющих точки, сортированные по фазе. Но тогда это должна быть сумма гипотенуз, а не их квадратов, т.е.

$$U(\Delta x_k, \Delta \phi_k) = \left((\Delta x_k / 2(x_{\max} - x_{\min}))^2 + (\Delta \phi_k)^2 \right)^{1/2}.$$

Можно вместо $2(x_{\max} - x_{\min})$ делить на $\alpha \sigma_0$, где α константа, значение которой можно подбирать различным для разных типов сигнала.

Еще одной из модификаций является метод Renson (1978), согласно которому находится сумма не квадратов гипотенуз, а квадратов наклонов гипотенуз, т.е.

$$U(\Delta x_k, \Delta \phi_k) = (\Delta x_k^2) / (\Delta \phi_k)^2.$$

Поскольку для некоторых пробных частот разность фаз может равняться нулю, это может привести к компьютерной ошибке деления на ноль. К тому же очень растет рассеяние на периодограммах. Чтобы избежать этого, Ренсон предложил ввести дополнительный параметр и включить его в виде

$$U(\Delta x_k, \Delta \phi_k) = (\Delta x_k^2) / ((\Delta \phi_k)^2 + \varepsilon^2)$$

или

$$U(\Delta x_k, \Delta \phi_k) = (\Delta x_k^2) / (\Delta \phi_k + \varepsilon)^2.$$

От выбора параметра ε будет зависеть периодограмма, но при малых значениях, как упоминалось, будет достаточно большое рассеяние на периодограмме, а при больших $\varepsilon \gg 1/N$ тест-функция будет практически пропорциональна тест-функции Лафлера-Кинмана-Холопова. Поэтому, если применять метод Ренсона, то имеет смысл выбирать $\varepsilon \sim 1/N$. Тщательный сравнительный анализ статистических свойств различных "непараметрических" методов был опубликован в статье Андронova и Чинарова в журнале «Кинематика и физика небесных тел» (1997 год, том 13, с. 67-80), поэтому в этой популярной статье приведены только простейшие и важнейшие формулы. На рис.3-5 этой статьи приведены масштабируемые периодограммы, рассчитанные для разных модификаций. Для нашего модельного сигнала все они показывают похожую структуру - шумовую дорожку для "неправильных частот" и минимумы у периодов $P_j = jP_0$.

Обобщение непараметрических методов

Несмотря на хорошую эффективность методов этого класса, они также имеют определенные недостатки. В-первых, периодограмма является прерывистой, поэтому за счет флуктуаций можно выбрать локальный минимум, что будет достаточно далеко как от минимума "средней периодограммы", так и от "настоящего" значения периода. Во-вторых, невозможно определить точность найденного периода (частоты), и можно только догадываться, что оптимальное значение периода находится в относительно широком минимуме на периодограмме. В-третьих, найденный период позволяет рассчитать фазовую кривую, но ее форму и соответственно начальную эпоху максимума или минимума все равно необходимо рассчитывать с помощью других (параметрических) методов. И четвертое, процедура сортировки наблюдений по фазе для каждой пробной частоты занимает значительно больше компьютерного времени, чем метод наименьших квадратов в "параметрических" методах. Первый недостаток можно уменьшить, если учитывать для каждого наблюдения не только две ближайшие по фазе точки, а все точки на фазовой кривой, только с дополнительным весом, зависят

от относительного расстояния между этими точками. Пельт (1980) записал общее выражение для тест-функции как

$$\Phi(f) = \sum_{kl=1}^N W(\phi_l - \phi_k) \cdot (x_l - x_k)^2 = \\ = 2 \sum_{k=2}^N \sum_{l=1}^{k-1} W(\phi_l - \phi_k) \cdot (x_l - x_k)^2,$$

где функция $W(x_l - x_k)$ называется "структурной" функцией. Например, можно находить квадраты разности со всеми точками, по фазе не отличающимися больше, чем на $\Delta \phi_+$ (например, 0,1). В этом случае $W(z) = 1$, если $z < \Delta \phi_+$, и 0, если $z > \Delta \phi_+$. Вместо этого "прямоугольного" фильтра можно предложить другие фильтры, отличные от нуля при $z < \Delta \phi_+$, например, "треугольный"

$$W(z) = 1 - |z| / \Delta \phi_+,$$

«колоколообразный»

$$W(z) = \left(1 - (z / \Delta \phi_+)^2 \right)^2,$$

или «косинусоидальный»

$$W(z) = \left(1 + \cos(\pi z / \Delta \phi_+) \right) / 2.$$

Последние два фильтра обеспечивают непрерывность не только самой тест-функции, но и ее производной.

Впрочем, если бы существовал «идеальный» метод периодограммного анализа, то все бы им и пользовались. Зачем были бы нужны другие методы? Но Природа такова, что в ней существуют очень разнообразные процессы, вызывающие разнообразие их проявлений. Поэтому исследователи должны иметь различные методы исследований, дополняющие друг друга. К одному из таких методов относятся рассмотренные «параметрические» и «непараметрические» методы периодограммного анализа.

Литература

1. Андронов И. Л. Визуальные и фотографические наблюдения переменных звезд. Одесса: ОГУ. 1991. 84с.
2. Андронов И. Л., Чинарова Л. Л. О статистических свойствах тест-функций непараметрических методов периодограммного анализа. // Кинематика и физика небесных тел.-1997. т. 13. с. 67-80.
3. Пельт Я. Частотный анализ астрономических временных рядов. Таллинн: Валгус - 1980.
4. Тербиж В. Ю. Анализ временных рядов в астрофизике. - М.: Наука.1992.-392с.
5. Тербиж В.Ю. Введение в статистическую теорию обратных задач - М.: Физматлит., 2005.- 376 с.
6. Холопов П. Н. Определение периодов изменения блеска переменных звезд с помощью электронных цифровых вычислительных машин // кн.: Методы исследования переменных звезд / Под ред. В. Б. Никонова. - М.: Наука.1971. - С. 307-329.

Дополнительная литература

1. Andronov, I. L. Phenomenological modeling of the light curves of algol-type eclipsing binary stars, 2012Ap....55..536A
2. Andronov, I. L. Advanced Methods for Determination of Arguments of Characteristic Events, 2005ASPC..335...37A
3. Andronov, I. L. Multiperiodic versus noise variations: mathematical methods, 2003ASPC..292..391A

Бесплатные программы для периодограммного анализа

1. Андронов И.Л., Бакланов А.В., MCV - Multi-Column Viewer, <http://uavso.pochta.ru/mcv/>
2. Бреус В.В. "VSCalc" - «Калькулятор переменных звезд», <http://uavso.org.ua/breus>

Иван Леонидович Андронов,

Одесский национальный морской университет (ОНМУ), доктор физико-математических наук, профессор, заведующий кафедрой «Высшая и прикладная математика»

Специально для журнала «Небосвод»