

УДК 621.371.31:550.338.2

ОПТИМИЗАЦИЯ ПОСТРОЕНИЯ ГИСТОГРАММ И СПЕКТРОГРАММ

А. Г. Вологдин, В. Д. Гусев

(кафедра физики атмосферы)

Для задач построения гистограмм (экспериментальных функций распределения вероятности) и спектрограмм (экспериментальных спектральных плотностей) решен вопрос минимизации суммарной ошибки, состоящей из статистической ошибки и ошибки аппроксимации. Показано, что относительная суммарная ошибка может быть существенно уменьшена при оптимальном выборе интервала группировки. Для некоторых видов распределений вероятности приведены конкретные результаты расчетов.

В практике обработки непрерывных случайных процессов $t(x)$ при построении экспериментальных дифференциальных и интегральных функций распределения вероятности наряду с главным вопросом о выборе числа наблюдений N (объема выборки), который определяет статистические ошибки, встает вопрос о длине интервала группировки Δx . Если длины Δx взять малыми, то будут доминировать статистические ошибки, если же длины Δx слишком велики, то будут утеряны детали функций распределения вероятности, т. е. увеличатся ошибки аппроксимации. Отметим, что в литературе относительно выбора интервала группировки Δx встречаются лишь некоторые, следующие из опыта, соображения (см., напр., [1]). На наш взгляд, задача выбора длины Δx должна решаться путем оптимизации.

В настоящей работе ставится задача анализа суммарной ошибки, состоящей из статистической ошибки и ошибки аппроксимации. Подчеркнем, что статистическая ошибка для функции распределения вероятности зависит как от объема выборки N , так и от интервала группировки Δx [2]. Следовательно, при решении вопроса оптимальности интервала группировки Δx необходимо учитывать суммарную ошибку.

Рассмотрим оценку $\hat{W}(x)$ плотности вероятности $W(x)$ случайной величины x . Точность оценки $\hat{W}(x)$ характеризуется средним квадратом ошибки

$$\sigma^2 = \langle [\hat{W}(x) - W(x)]^2 \rangle = \sigma_{\hat{W}}^2 + s_{\hat{W}}^2, \quad (1)$$

где

$$\sigma_{\hat{W}}^2 = \langle [\hat{W}(x) - \langle \hat{W}(x) \rangle]^2 \rangle$$

является дисперсией оценки, а

$$s_{\hat{W}} = \langle \hat{W}(x) - W(x) \rangle$$

представляет собой ее смещение. В [2] при объеме выборки N для дисперсии оценки приведено приближенное выражение

$$\sigma_{\hat{W}}^2 + W(x)/N \Delta x, \quad (N \Delta x \gg 1).$$

Там же указывается, что смещение оценки

$$s_{\hat{W}} \simeq \Delta x^2 W''(x)/24$$

в области максимальных значений плотности вероятности $W(x)$ мало и им в формуле (1) можно, как правило, пренебречь. Это обстоятель-

ство проверялось ниже в каждом конкретном случае, и ошибка за счет смещения не превышала 1%.

Таким образом, учитывая малость смещения, для относительной статистической ошибки имеем

$$\delta_1 = \sigma/W(x) \simeq \sigma_{\hat{W}}/W(x) = [NW(x)\Delta x]^{-1/2}.$$

Относительная ошибка аппроксимации плотности вероятности $W(x)$ определяется формулой

$$\delta_2 = \Delta W(x)/W(x),$$

причем δ_1 и δ_2 статистически независимы. В области, где $W'(x) \neq 0$,

$$\delta_2 = W'(x)\Delta x/W(x).$$

Для определения оптимального интервала группировки Δx_{opt} на квадрат суммарной ошибки

$$\delta^2 = \delta_1^2 + \delta_2^2 = [NW(x)\Delta x]^{-1} + [W'(x)\Delta x/W(x)]^2 \quad (2)$$

накладываем условие $\delta^2 = \min$. Отсюда получаем уравнение для Δx_{opt} :

$$\delta_{\Delta x}^2 = \{[NW(x)\Delta x]^{-1} + [W'(x)\Delta x/W(x)]^2\}'_{\Delta x} = 0,$$

решение которого дает

$$\Delta x_{\text{opt}} = [2NW'(x)W^{-1}(x)]^{-1/3}. \quad (3)$$

Рассмотрим некоторые примеры.

1. Нормальный закон распределения вероятности

$$W(x) = \exp\{-x^2/2\}/2^{-1/2}\pi^{-1/2}.$$

В этом случае получим из (3) и (2)

$$\Delta x_{\text{opt}} = [2Nx^2W(x)]^{-1/3},$$

$$\delta_{\min} = \{[NW(x)\Delta x_{\text{opt}}]^{-1} + (x\Delta x_{\text{opt}})^2\}^{1/2}.$$

Расчеты по этим формулам сведены в табл. 1, где приведены данные при объеме выборки $N=10^2, 10^3$ для значений плотности вероятности, близких к максимальному значению W_m . Последняя колонка таблицы δ_* содержит процент отклонения ошибки при неоптимальном выборе, который для примера выбран равным $\Delta x = \Delta x_{\text{opt}}/2$, по отношению к δ_{\min} , т. е.

$$\delta_* = (\delta_{\text{opt}} - \delta_{\min}) \delta_{\min}^{-1} \cdot 100\%.$$

Содержание других таблиц аналогично по смыслу таблице 1.

Таблица 1

$W(x)$	$ x $	Δx_{opt}		$\delta_{\min}, \%$		$\delta_*, \%$	
		$N=10^2$	$N=10^3$	$N=10^2$	$N=10^3$	$N=10^2$	$N=10^3$
0,95 W_m	0,32	0,49	0,23	27	13	33	15
0,90 W_m	0,46	0,40	0,18	32	14	19	29
0,70 W_m	0,84	0,29	0,13	42	19	21	26
0,50 W_m	1,18	0,26	0,12	53	24	21	25

2. Плотность вероятности Рэля

$$W(x) = x \exp\{-x^2/2\}, \quad x \geq 0.$$

Из (3) и (2) следует, что

$$\Delta x_{\text{opt}} = [2N(x - x^{-1})^2 W(x)]^{-1/3},$$

$$\delta_{\text{min}} = \{[(x - x^{-1}) \Delta x_{\text{opt}}]^2 + [N \Delta x_{\text{opt}} W(x)]^{-1}\}^{1/2}.$$

Таблица 2 содержит данные, рассчитанные по этим формулам.

Таблица 2

W(x)	x	Δx_{opt}		$\delta_{\text{min}}, \%$		$\delta_*, \%$	
		N=10 ²	N=10 ³	N=10 ²	N=10 ³	N=10 ²	N=10 ³
0,95 W _m	0,78	0,33	0,15	28	13	18	23
	1,23	0,36	0,17	27	13	19	15
0,90 W _m	0,69	0,23	0,11	33	15	24	27
	1,32	0,30	0,14	30	14	20	21
0,7 W _m	0,47	0,16	0,08	46	21	22	24
	1,65	0,22	0,10	40	18	20	22
0,50 W _m	0,32	0,13	0,06	62	28	19	21
	1,92	0,20	0,09	49	22	20	27

3. Плотность вероятности

$$W(x) = 2^{-1} (1 + x^2)^{-3/2}.$$

Данная формула описывает плотность вероятности производной фазы квазигармонического сигнала со случайной начальной фазой, имеющей нормальный закон распределения вероятности [3].

Как и выше, из (3) и (2) имеем

$$\Delta x_{\text{opt}} = [18N(x + x^{-1})^{-2} W(x)]^{-1/3},$$

$$\delta_{\text{min}} = \{[N \Delta x_{\text{opt}} W(x)]^{-1} + [3x(1 + x^2)^{-1} \Delta x_{\text{opt}}]^2\}^{1/2}.$$

Данные для этой плотности вероятности приведены в табл. 3.

Таблица 3

W(x)	x	Δx_{opt}		$\delta_{\text{min}}, \%$		$\delta_*, \%$	
		N=10 ²	N=10 ³	N=10 ²	N=10 ³	N=10 ²	N=10 ³
0,95 W _m	0,19	0,33	0,15	30	14	23	21
0,90 W _m	0,27	0,26	0,12	35	16	20	25
0,70 W _m	0,52	0,21	0,10	44	20	23	25
0,50 W _m	0,77	0,23	0,11	53	24	15	21

Анализ данных, содержащихся в таблицах, позволяет сформулировать некоторые общие выводы.

Относительная ошибка при построении гистограмм при заданном объеме выборки может быть существенно уменьшена при оптимальном выборе интервала группировки Δx . Как увеличение, так и уменьшение Δx по отношению к Δx_{opt} ведет к увеличению суммарной ошибки. Оптимальный интервал Δx_{opt} уменьшается при увеличении объема

выборки N . Для разных значений плотности вероятности $W(x)$ интервал Δx_{opt} — разный, а относительная ошибка δ_{min} увеличивается с уменьшением $W(x)$.

В заключение отметим, что при экспериментальной оценке спектральных плотностей $G(\omega)$, т. е. при построении спектрограмм, ситуация сходна, а аналогом длины Δx является величина спектрального окна $\Delta\omega$. Следовательно, задача оптимального выбора окна $\Delta\omega_{\text{opt}}$ решается тем же путем. Формулы (1), (2) при замене $W(x)$ на $G(\omega)$, N на T (длительность интервала наблюдения) и Δx на $\Delta\omega$ являются решением задачи оптимального выбора спектральных окон $\Delta\omega_{\text{opt}}$.

ЛИТЕРАТУРА

1. Арлей Н., Бух К. Введение в теорию вероятности и математическую статистику. М., 1951.
2. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. М., 1989.
3. Тихонов В. И. Статистическая радиотехника. М., 1986.

Поступила в редакцию
14.12.94

ВЕСТН. МОСК. УН-ТА. СЕР. 3. ФИЗИКА. АСТРОНОМИЯ. 1995. Т. 36, № 4

ОПТИКА И СПЕКТРОСКОПИЯ

УДК 532.74; 535.37

ТЕМПЕРАТУРНЫЕ ИЗМЕНЕНИЯ СПЕКТРАЛЬНЫХ ХАРАКТЕРИСТИК ВОДНЫХ РАСТВОРОВ ГЛИЦИЛТРИПТОФАНА

Л. И. Иваньян, Л. В. Левшин, Б. Д. Рыжиков

(кафедра общей физики)

Показано, что наблюдаемые изменения интенсивности комбинационного рассеяния света и люминесценции при изменении температуры водных растворов глицилтриптофана зависят от предыстории раствора. При изменении температуры происходит деструкция раствора, при этом для установления новой равновесной структуры при фиксированной температуре требуется более 14 часов.

В последнее время появился ряд научных публикаций, посвященных изучению температурных зависимостей оптических спектров воды. Особый интерес представляют работы по изучению температурных изменений спектров комбинационного рассеяния света (КРС) воды и водных растворов. Так, в работах [1, 2] проводится анализ зависимости формы линии КРС от температуры. В работах [3, 4] обсуждается зависимость интенсивности КРС от температуры и делается вывод о сохранении динамического равновесия свободных и участвующих в образовании Н-связей ОН-групп в изучаемых интервалах температур. Это подтверждается данными, изложенными в [5], где установлено существование изобестических точек, для которых интенсивность КРС не зависит от температуры. В работе [6] с использованием формализма поляризуемости связей в изолированных молекулах моделируются различные структуры жидкой воды, а в [7] сложная структура спектра КРС связывается со структурными особенностями воды на основе применения дуэструктурной модели воды.

Однако до сих пор не существует единой интерпретации полученных разными авторами результатов.