УДК 53.088.3; 53.088.7

Ю.К. ТАРАНЕНКО, О.Ю. ОЛЕЙНИК
Украинский державний химико-технологический университет, г. Днепр

РАЗРАБОТКА МОДЕЛИ ОБРАБОТКИ ДАНЫХ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ПРЕОДОЛЕНИЯ АПРИОРНОЙ НЕОПРЕДЕЛЕННОСТИ

В статье приведено описание модели для обработки данных при условии априорной неопределенности. Обосновано преимущество использование байесовского подхода в решении задачи параметрической априорной неопределенности. Приведено вид практической реализации в программной среде Python.

Ключевые слова: *модель, вероятность, байесовский подход, результаты измерения, дисперсия.итерационные методы Монте-Карло.*

YU.K. TARANENKO, O.YU. OLIYNYK
State Higher Educational Institution «Ukrainian State Chemical Technology University»

THE ARTICLE DESCRIBES A MODEL FOR PROCESSING DATA UNDER THE CONDITION OF A PRIORI UNCERTAINTY. THE ADVANTAGE OF USING THE BAYESIAN APPROACH IN SOLVING THE PROBLEM OF PARAMETRIC A PRIORI UNCERTAINTY IS SUBSTANTIATED. THE FORM OF PRACTICAL IMPLEMENTATION IN THE PYTHON SOFTWARE ENVIRONMENT IS GIVEN.

The article describes a model for processing data under the condition of a priori uncertainty. The advantage of using the Bayesian approach in solving the problem of parametric a priori uncertainty is substantiated. The form of practical implementation in the Python software environment is given.

Key words: model, probability, Bayesian approach, measurement results, variance. Iterative Monte Carlo methods.

Введение

В информационно-измерительных системах, работе систем диагностики технического состояния объекта возникает ситуация, когда неизвестен факт не только наличия и отсутствия сигнала на входе приемника, но и неизвестны характеристики помех и сигнала. Возникает проблема априорной неопределенности или неизвестности [1]. Выделяют параметрическую априорную неопределенность и непараметрическую неопределенность. Первая связана с ситуацией, когда при известном законе распределения вероятностей сигнала и помехи неизвестны значения параметров этого закона распределения. Вторая‒ когда неизвестен закон распределения сигнала и помехи. Возникает нехватка априорных данных, что не дает возможности установить связь наблюдаемых величин с условным риском. При дополнительной обработке входной информации нужно восстановить соответствие между ожидаемыми потерями и этой информацией. Такой процесс называют адаптацией. Чаще всего эти правила формируют в рамках так называемого адаптивного байесова подхода, основной особенностью которого является замена неизвестных параметров, характеристик или законов распределения помех их состоятельными оценками. При некоторых ограничениях эти оценки становятся оценками максимального правдоподобия [2,3].

Анализ последних публикаций и постановка проблемы

На сегодняшний день основными методами преодоления априорной неопределенности являются:

- использование адаптации к неизвестным или меняющимся параметрам помехи (создание адаптивных параметрических систем);
- создание устройств обнаружения, нечувствительных к виду закона распределения вероятностей помех (создание адаптивно-непараметрических, или инвариантных систем);
- использование систем обнаружения сигнала, стабильно работающих и незначительно теряющих свои свойства при изменении законов распределения вероятностей помех(создание робастных систем) [1].

Так, например, в работе [4]приведено описание робастных обнаружителей, основанных на минимаксном правиле Неймана - Пирсона. Критерий различения гипотез наличия и отсутствия сигнала базируется на отношении правдоподобия и минимизирует максимальный риск пропуска сигнала при фиксированном риске ложной тревоги.

Однако, главным недостатком перечисленных методов преодоления априорной неопределенности являются то, что они аппаратные и их применение заключается в разработке технического средства специальной конструкции, реализующего конкретную задачу.

Байесовские подходы, известные исследователям с 1930 гг., разработаны с целью решения проблемы статистического анализа поведения процессов и систем разной природы на основе теоремы Байеса, которая использует соотношения между вероятностями событий разного характера и спецификации любого события на необходимом уровне [5]. Основным аспектом, который затруднял широкое распространение метода, была необходимость значительных вычислительных затрат, связанных с численным интегрированием в многомерных пространствах, знаний статистики. Современные информационные технологии устраняют трудности связанные со сложностью вычислений.

Байесовский подход уже был применен при решении многих задач в разных областях науки и техники [6].Отличительной чертой байесовского метода является использование априорной информации относительно параметров модели. Такая информация выражается в виде априорной вероятности или функции плотности вероятности. Затем начальные вероятности «пересматриваются», с помощью выборочных данных, которые находят свое отображение в виде апостериорного распределения оценок

параметров или переменных модели.

В [7] описан байесовский подход к обработкеданных, суть которого сводиться к тому, что заранее задаётся вид исследуемой зависимости (от некоторого аргумента) с вектором её параметров. Описанные подходы к обработке экспериментальных данных сводятся к статистическим методам, с учетом выбранных предположений о представительности выборки, о нормальности распределения погрешностей измерения, о статистической однородности погрешности многократных измерений и о несмещённости замеров, а также на предположение, что значенияаргумента известны точно[7].

Цель и задачи исследования

Целью данной работы: раскрыть существующие подходы исследований с использованием байесовских методов обработки данных и применить их при создании модели обработки данных гипотетического эксперимента. Для достижения поставленной цели в задаче необходимо решить следующие задачи:

разработать модель обработки данных гипотетического эксперимента и реализовать ее в в программной среде Python;

для данных полученных в ходе гипотетического эксперимента, поступающих с шумом, дисперсия которого неизвестна необходимо определить коэффициенты линейной зависимости.

оценить результаты обработки экспериментальных данных с использованием байесовских методов путем сравнения с результатами обработки данных известными методами.

Создание модели в программной среде Python

Модель была реализованав программной средеPython3.4. (модули: numpy, matplotlib, pymc).Python является простым и, в то же время, мощным интерпретируемым объектно-ориентированным языком программирования. Он предоставляет структуры данных высокого уровня, имеет изящный синтаксис и использует динамический контроль типов, что делает его идеальным языком для быстрого написания различных приложений, работающих на большинстве распространенных платформ [8]. Кроме того, Python является свободно доступным программным продуктом, что дает возможность широкого использования результатов разработки.

Суть полученной модели лежит в следующем: создается модель сбора данных, которая зависит от измеряемого параметра. В данном случае цель интерполяция данных с помощью прямой вида $y = ax + b$ (делаем предположение, что все данные имеют линейную зависимость с наложенным на нее гауссовым шумом с известной дисперсией). Тогда a и b – это выступают в роли параметров, вероятные значения которых необходимо определить, а функция правдоподобия – гауссова функция со средним, заданным уравнением прямой, и данной дисперсией[9].Априорная вероятность включает в себя информацию, которую мы знаем до проведения анализа. Например, мы точно знаем, что прямая должна иметь положительный наклон, или, что значение в точке пересечения с осью y должно быть положительным, – все это учитывает предложенная модель. При задании априорного распределения необходимо указать первый аргумент, а также указать границы распределения (второй и третий аргументы):

```
sigma = pymc.Uniform('sigma', 0., 100.)
a     = pymc.Uniform('a', 0., 20.)
b     = pymc.Uniform('b', 0., 20.)
```

В модуле pymc определяем модель как детерминистическую:

```
@pymc.deterministic(plot=False)
deflinear_fit(a=a, b=b, x=x):
return a*x + b
```

Функция правоправдоподобия выглядит следующим образом:

```
y = pymc.Normal('y', mu=linear_fit, tau=1.0/sigma**2, value=data,
observed=True)
```

Листинг созданной модели имеет вид:

```
importnumpy
importpymc
#Generatedatawithnoise
number_points     = 20
true_coefficients = [8.2, .5.4]
x                 = numpy.linspace(0, 10, number_points)
noise             = numpy.random.normal(size = number_points)
data              = true_coefficients[0]*x + true_coefficients[1] + noise
#PRIORs:
#assigmaisunknownthenwedefineitas a parameter:
sigma = pymc.Uniform('sigma', 0., 100.)
#fittingtheline y = a*x+b, hencethecoefficientareparameters:
a     = pymc.Uniform('a', 0., 20.)
b     = pymc.Uniform('b', 0., 20.)
#definethemodel: if a, b and x aregiventhereturnvalueisdetermined,
hencethemodelisdeterministic:
@pymc.deterministic(plot=False)
deflinear_fit(a=a, b=b, x=x):
return a*x + b
#LIKELIHOOD
```

```
        #normallikelihoodwithobserveddata (withnoise), modelvalueandsigma
        y = pymc.Normal('y', mu=linear_fit, tau=1.0/sigma**2, value=data,
observed=True)
```

Генерирование случайных величин с заданным распределением — актуальный подход, который срабатывает в условиях, когда существуют асимптотические аналитические результаты для свойств оценок и их статистических распределений, но неизвестно, какие свойства будут иметь оценки при малых выборках данных. Модуль pymc реализует алгоритм итерационных методов Монте-Карло для марковских цепей (MCMC)[10]. Модель генерирует (делает выборку) значений a, b и sigma, а затем выполняет расчет p(Data | a, b, sigma). В результате получается ряд значений, который представляет собой выборку из апостериорного распределения:

```
        D = pymc.MCMC(model, db = 'pickle')
        D.sample(iter = 10000, burn = 1000)
```

D.sample принимает два аргумента (можно задать и больше) - количество итераций, burn-in- это количество первых итераций, которые не учитываются. Для получения сравнительных характеристик используемого байесовского подхода в модели предусмотрим решение задачи методом наименьших квадратов:

```
        chisq_result = polyfit(model.x, model.data, 1)
```

Листинг подпрограммы расчета имеет вид:

```
        fromnumpy             import polyfit
        frommatplotlib.pyplot import figure, plot, show, legend
        importpymc
        import model
        #Define MCMC:
        D = pymc.MCMC(model, db = 'pickle')
        #Sample MCMC: 10000 iterations, burn-in period is 1000
        D.sample(iter = 10000, burn = 1000)
        #compute chi-squared fitting for comparison:
        chisq_result = polyfit(model.x, model.data, 1)
        #print the results:
        print "nnResult of chi-square result: a= %f, b= %f" % (chisq_result[0],
chisq_result[1])
        print "nResult of Bayesian analysis: a= %f, b= %f" % (D.a.value, D.b.value)
        print "nThe real coefficients are:   a= %f, b= %fn"
%(model.true_coefficients[0], model.true_coefficients[1])
        #plot graphs from MCMC:
        pymc.Matplot.plot(D)
        #plot noised data, true line and two fitted lines (bayes and chi-squared):
        figure()
        plot(model.x, model.data, marker='+', linestyle='')
        plot(model.x, D.a.value * model.x + D.b.value, color='g', label='Bayes')
        plot(model.x, chisq_result[0] * model.x + chisq_result[1], color='r',
label='Chi-squared')
        plot(model.x, model.true_coefficients[0] * model.x +
model.true_coefficients[1], color='k', label='Data')
        legend()
        show()
```

Для проверки адекватности созданной модели и сопоставления результатов предусмотрено расчет значения реальных коэффициентов линейной зависимости, коэффициентов рассчитанных по методу наименьших квадратов и коэффициенты рассчитанные с использованием байесовских методов. Расчет проводился для значения коэффициентов a=8,2; b=5,4.Результаты расчета выводятся в отдельном окне:

```
        Result of chi-square result: a= 8.271636, b= 5.229759
        Result of Bayesian analysis: a= 8.255764, b= 5.403548
        The real coefficients are:   a= 8.200000, b= 5.400000
```

Полученные данные позволяют говорить о возможности применения модели, для случаев, года априорные данные отсутствуют.

На рис. 1. представлена графическая часть результата работы модели.

Окно графической части представляет собой гистограмму апостериорного распределения для параметров (в данном случае приведен параметр a). Слева от гистограммы приводиться подтверждение сходимости MCMC и след выборки после применения байесовского метода.

Выводы

Описанная модель для гипотетических данных, поступающих с шумом, дисперсия которого неизвестна,создаёт плодотворную базу для дальнейших исследований в области обработки сильно зашумленных результатов измерения. Данные [9] свидетельствуют о успешном применении модели для обработки результатов измерения, полученных при макетных испытаниях средства измерения.

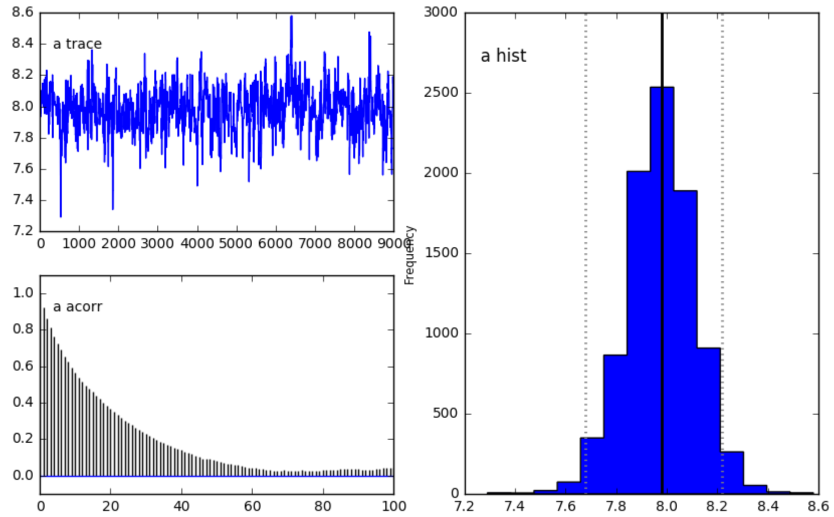

Рис.1.- Результаты работы модели

Литература

1. Бакулев П.А. Радиолокационные системы. Учебник для вузов. — М.: Радиотехника, 2004, - 320 с.
2. Марчук В.И. Повышение достоверности первичной обработки результатов измерений[Текст] / В.И. Марчук // Измерительная техника, −2003.−№12. −С. 3–5.
3. Марчук В.И. Новый подход обработки результатов измерений при априорной неопределенности. [Текст] / В.И. Марчук, А.И. Шерстобитов// Труды РНТОРЭС им. Попова. Серия: Цифровая обработка и её применение.–М.:, Вып. 7. т.1,2005. – С.48 – 51.
4. Неволин В.И. Робастная обработка сигналов в ультразвуковой дефектоскопи [Текст] / Неволин В.И., Беленьков В.В. // Контроль. Диагностика.М.:− 2011. –Т.№10.−С.37-45.
5. Беседин А.Н. Обработка случайных сигналов и процессов / А.Н. Беседин, А.А. Зеленский, Г.П. Кулемин, В.В. Лукин. – Учеб.пособие. – Харьков: Нац. аэрокосм. ун-т «Харьк. авиац. ин-т», 2005. – 469 с.
6. Berkovskii N.A. Error of calculating the optimal Bayesian estimate using the Monte Carlo method in nonlinear problems[Text] /N.A. Berkovskii, O.A. Stepanov // Journal of Computer and Systems Sciences International. 2013. T. 52. № 3. P. 342-353.
7. Pilz J. Some thoughts on the present position in Bayesian statistics// Mathematical Research. — 1990. — Vol.68. — P. 70-82.
8. Россум Г.. Язык программирования Python. −2001 — 454 с.
9. Тараненко Ю.К. Применение байесовских методов при обработке сильно зашумленных результатов измерений [Текст] / Ю.К. Тараненко, О.Ю. Олейник // Вимірювальна та обчислювальна техніка в технологічних процесах.− 2017. –Т.№1.− С. 205-210.
10. Cheung SH, Beck JL. Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters/ Cheung SH, Beck JL.// J. Eng. Mech.2009.– 135, P. 243–255.

References

1. Bakulev P.A. Radiolokatsionnyye sistemy. Uchebnik dly avuzov. — M.: Radiotekhnika, 2004, - 320 s.
2. Marchuk V.I. Povysheniye dostovernosti pervichnoy obrabotki rezul'tatov izmereniy [Tekst] / V.I. Marchuk // Izmeritel'naya tekhnika, −2003.− №12. − S. 3–5.
3. Marchuk V.I. Novyy podkhod obrabotki rezul'tatov izmereniy pri apriornoy neopredelennosti. [Tekst] / V.I. Marchuk, A.I. Sherstobitov // Trudy RNTORES im. Popova.Seriya: Tsifrovaya obrabotka i yeyo primeneniye.–M.:,Vyp. 7. t.1, 2005. – S.48 – 51.
4. Nevolin V.I. Robastnaya obrabotka signalov v ul'trazvukovoy defektoskopi [Tekst] / Nevolin V.I., Belen'kov V.V. // Kontrol'.Diagnostika. M.:− 2011. –T.№10.− S.37-45/
5. Besedin A.N. Obrabotka sluchaynykh signalov i protsessov / A.N. Besedin, A.A. Zelenskiy, G.P. Kulemin, V.V. Lukin. – Ucheb. posobiye. – Khar'kov: Nats. aerokosm. un-t «Khar'k. aviats. in-t», 2005. – 469 s.
6. Berkovskii N.A.Error of calculating the optimal Bayesian estimate using the Monte Carlo method in nonlinear problems [Text] / N.A. Berkovskii, O.A. Stepanov // Journal of Computer and Systems Sciences International. 2013. T. 52. № 3. P. 342-353.
7. Pilz J. Some thoughts on the present position in Bayesian statistics// Mathematical Research. — 1990. — Vol.68. — P. 70-82.
8. G. Rossum. YAzykprogrammirovaniya Python.−2001 — 454 c.
9. Taranenko Yu K. Primeneniye bayyesovskikh metodov pri obrabotke sil'nozashumlennykh rezul'tatov izmereniy [Tekst] / YU.K.Taranenko, O.YU. Oleynik // Vimíryuval'na ta obchislyuval'na tekhníka v tekhnologíchnikh protsesakh.− 2017. –T.№1.− S. 205-210..
10. Cheung SH, Beck JL. Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters/ Cheung SH, Beck JL. // J. Eng. Mech.2009.– 135, P. 243–255.

Рецензія/Peer review : 13.4.2017 р. Надрукована/Printed :22.6.2017 р.
Стаття рецензована редакційною колегією