КОНФЕРЕНЦИЯ D

ФИЗИКА ТРОПОСФЕРЫ

РОБАСТНЫЕ НЕПАРАМЕТРИЧЕСКИЕ ДАТЧИКИ СЛУЧАЙНЫХ ВЕЛИЧИН ДЛЯ ОБРАБОТКИ МИНИ-СОДАРНЫХ ИЗМЕРЕНИЙ

Симахин В.А.,[1] Шаманаева Л.Г,[2,3] Маер А.В.[1]

[1]Курганский государственный университет

[2]Институт оптики атмосферы им. В.Е. Зуева СО РАН

[3]Национальный исследовательский Томский государственный университет

e-mail: sva_full@mail.ru; sima@iao.ru.

В докладе предложены бутстреп-процедуры обработки данных физических экспериментов при решении полунепараметрических задач. Предложен алгоритм построения робастного непараметрического датчика случайных величин, который использован для создания алгоритмов бутстреп-процедур. Полунепараметрические бутстреп-процедуры использованы для обработки результатов мини-содарных измерений компонентов скорости ветра в пограничном слое атмосферы.

С появлением высокопроизводительных и доступных средств вычислительной техники появилась возможность исследования сложных систем и решения задач, недоступных для аналитического рассмотрения, с помощью методов Монте-Карло [1, 2]. Незнание вида распределения и наличие выбросов привели к созданию непараметрических и робастных процедур принятия решений [3, 4]. Желание узнать свойства статистических процедур при конечных объемах выборки и неизвестном распределении привело к разработке компьютерного метода исследования распределения статистик вероятностных распределений, основанного на многократной генерации выборок методом Монте-Карло на базе имеющейся выборки – бутстрепа, основой которого являются непараметрические датчики случайных величин [2]. Вопросы обработки данных для полунепараметрических задач в *условиях ограниченной выборки, неизвестного распределения и наличия выбросов* давно привлекают внимание исследователей. Традиционный подход, выработанный экспериментаторами, связан с *очисткой* выборки от выбросов и дальнейшей обработкой результатов экспериментов методами параметрической и непараметрической статистик и бутстрепа. В данной работе предложена бутстреп-процедура обработки данных с использованием полунепараметрических датчиков случайных величин.

Пусть X – случайная величина с функцией распределения (ф.р.) $F\left(x,\vec{\theta}\right)\subset\Omega$, где Ω – класс распределений Тьюки $F\left(x,\vec{\theta}\right)=(1-\varepsilon)G\left(x,\vec{\theta}\right)+\varepsilon H(x)$, $G\left(x,\vec{\theta}\right)$ – априорное распределение, $H(x)$ и ε – распределение и доля выбросов, $f\left(x,\vec{\theta}\right), g\left(x,\vec{\theta}\right), h\left(x\right)$ – соответствующие плотности распределений, $\vec{\theta}=\left(\theta_1,\ldots,\theta_k\right)^T$ – вектор неизвестных параметров распределения. Для построения датчика случайных величин воспользуемся классическим методом генерирования случайных величин с помощью обратного преобразования $X=F^{-1}(U)$ в виде $x_i=F^{-1}(u_i), i=1,2,\ldots$, где $u_i, i=1,2,\ldots$ реализация равномерных на [0, 1] случайных величин [1]. Генерация выборочного значения x_i из распределения $F(x)$ сводится к нахождению квантиля распределения $F(x)$ уровня u_i. Метод получения непараметрического датчика заключается в следующем. Вместо неизвестной ф.р. $F(x)$, подставляют её непараметрическую оценку (ЭФР) $\bar{F}_N(x)$ и затем решают оценочное уравнение $X=\bar{F}_N^{-1}(U)$. Обозначим через X_p единственный квантиль распределения $F(x)$ уровня p $(0<p<1)$. Предположим, что уравнение $F\left(X_p,\vec{\theta}\right)=p$ имеет единственное решение. Требуется по выборке $\vec{X}_N=\left(x_1,\ldots,x_N\right)$ объема N независимых и одинаково распределенных случайных величин из $F\left(x,\vec{\theta}\right)\subset\Omega$ построить оценку X_{pN} для квантиля X_p. Задачи и методы нахождения оценок квантиля X_{pN} определяются априорной информацией о виде ф.р. $F\left(x,\vec{\theta}\right)\subset\Omega$. Рассмотрим полунепараметрическую задачу [3]. В этом случае функция распределения $F\left(x,\vec{\theta}\right)\subset\Omega_n$, при этом $\left\{G\left(x,\vec{\theta}\right)\subset\bar{\Omega}_n;[H(x)\subset\Omega_n,\varepsilon\neq 0]\right\}$, Ω_n – непараметрический класс функций распределения, $\bar{\Omega}_n$ – непараметрический класс функций распределения при некоторой дополнительной информации о $G\left(x,\vec{\theta}\right)$ – задачи робастной непараметрической статистики [3, 4]. Отметим, что вопрос идет о квантиле априорного распределения $G\left(X_p,\vec{\theta}\right)=p$ на фоне выбросов в распределении $F\left(X,\vec{\theta}\right)$ из которого и берется неоднородная выборка $\vec{X}_N=\left(x_1,\ldots,x_N\right)$. Данные задачи наиболее типичны и актуальны при обработке результатов физических экспериментов [1, 5]. В качестве состоятельной оценки X_{pN} квантиля X_p функции распределения $F\left(x\right)$ возьмем решение эмпирического уравнения

$$\hat{F}_N\left(X_p\right)=p \;\Rightarrow X_{pN}=\hat{F}_N^{-1}\left(p\right), \tag{1}$$

где $\hat{F}_N\left(x\right)$ – состоятельная и несмещенная (асимптотически несмещенная) оценка $F\left(x\right)$.

Для нахождения оценки X_{pN} обычно используются рекуррентные методы стохастической аппроксимации вида

$$X_{pN}\left[k+1\right]=X_{pN}\left[k\right]-\gamma\left[k\right]\left[\hat{F}_N\left(X_{pN}\left[k\right]\right)-p\right],\ k=0,1,\dots, \qquad (2)$$

где $\gamma\left[k\right]$ должны удовлетворять условиям: $\gamma\left[k\right]>0,\ \sum_{k=0}^{\infty}\gamma\left[k\right]=\infty,\ \sum_{k=0}^{\infty}\gamma^2\left[k\right]<\infty$. Обозначим через $\tilde{F}_N\left(x\right)$ робастную непараметрическую оценку $F\left(x\right)$. Для получения робастных непараметрических оценок квантиля $X_{pN}=\tilde{F}_N^{-1}\left(p\right)$, необходимы $\tilde{F}_N\left(x\right)$. Для случая $\left\{F(x)=G\left(x\right)\subset\Omega_n[\varepsilon=0]\right\}$, непараметрические оценки $X_{pN}=F_N^{-1}\left(p\right)$ для квантиля на порядковых статистиках широко известны [3, 4]. Полупараметрические высокоэффективные оценки функции распределения $\tilde{F}_N\left(x\right)$ и плотности $\tilde{f}_N(x)$ на основе взвешенного метода максимального правдоподобия (ВММП) рассматривались в [3], где для полупараметрической оценки квантиля X_{pN} на основе ВММП получено следующее оценочное уравнение:

$$\sum_{i=1}^{N}\sum_{j=1 j\neq i}^{N}(\mu_N-u_{ij})\cdot\tilde{f}_N^{-1}(x_i,\mu_N,h_N)\exp\left[-\frac{1}{2h_N^2}(\mu_N-u_{ij})^2\right]=p, \qquad (3)$$

где
$$\tilde{f}_N(x_i,X_{pN},h_N)=\frac{1}{N\sqrt{2\pi h_N^2}}\sum_{j=1}^{N}\left\{\exp\left[-\frac{1}{\sqrt{2}h_N^2}(X_{pN}-v_{ij})^2\right]\right\},$$

$$u_{ij}=\frac{y_i+y_j}{h_N},\ \mu_N=\tilde{F}_N\left(X_{pN}\right),\ y_i=C(X_{pN}-x_i),\ t_i=k((X_{pN}-x_i)\cdot h_N^{-1}),$$

$v_{ij}=\dfrac{t_i+t_j}{h_N}$, и $k(u)=\dfrac{1}{\sqrt{2\pi h_N^2}}\exp\left[-\dfrac{1}{2h_N^2}u^2\right]$ – нормальная ядерная функция. В данном случае получаем полунепараметрические адаптивные оценки ВММП в виде взвешенного среднего полусумм Уолша. Для нахождения l_{opt}, $X_{pN}(l_{opt})$, используются алгоритмы непараметрической адаптации, рассмотренные в [3].

Полученные оценки использовались для обработки данных допплеровского мини-содара AV4000. Рабочая частота содара 4900 Гц, длительность импульса излучения 60 мс, период повторения импульсов 4 с. Излучение последовательно посылалось в трех направлениях – вертикально вверх и под углами 14° к вертикали в двух взаимно ортогональных плоскостях. Анализировались данные измерений трех компонентов скорости ветра в 43 высотных стробах вертикальной протяженностью 5 м в диапазоне высот 5–200 м. Обрабатывались серии из N = 150 профилей, что обеспечивало усреднение за 10-минутный период измерения.

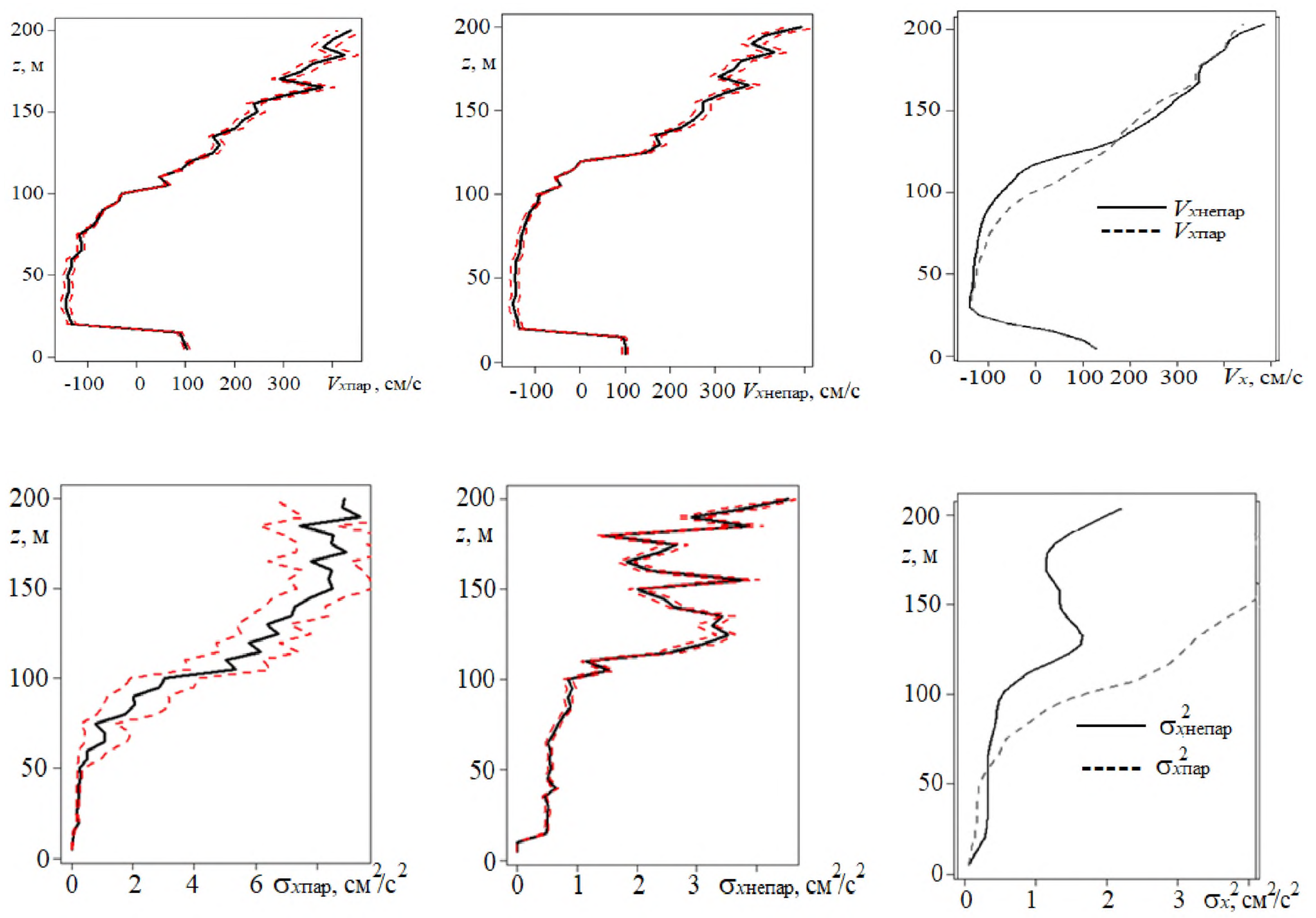

Рисунок 1 — Параметрические и полупараметрические оценки высотных профилей x-компонентов скорости ветра $V_{xпар}(z)$ и$V_{xнепар}(z)$ b их дисперсий $\sigma^2_{xпар}(z)$ и $\sigma^2_{xнепар}(z)$ с доверительными интервалами (красные пунктирные кривые) по результатам минисодарных измерений в ночные часы с 00:00 до 00:10 местного времени и их сравнение (справа на рисунке).

Из рисунка видно, что бутстреп процедура существенно уменьшает дисперсию полученной оценки. Результаты обработки мини-содарных измерений в АПС показывают, что стандартные методы обработки приводят к значительным смещениям и низкой эффективности оценок по сравнению с семипараметрическими оценками максимального правдоподобия.

Работа выполнена в рамках государственного задания II.12.1.2 ИОА СО РАН (рег. № проекта АААА-А17-117021310152-4).

1. *Ермаков С. М., Михайлов Г. А.* Курс статистического моделирования. М.: Наука, 1976. 320 с.

2. *Эфрон Б.* Нетрадиционные методы многомерного статистического анализа. М.: Финансы и статистика, 1998. 263 с.

3. *Симахин В. А.* Адаптивные оценки. Курган: КГУ, 2019. 240 с.

4. *Шуленин В.П.* Робастные методы математической статистики. Томск: Изд-во НТЛ, 2016. 260 с.

5. *Федоров В.А.* Измерение содаром "Волна-3" параметров радиальных компонент вектора скорости ветра // Оптика атмосферы и океана. 2003. Т. 16. № 02. С. 151-155.