Методические указания
по выполнению лабораторной работы № 4
по курсу «Обработка данных в точных измерительных системах»

Суть обработки данных заключается в получении количественных характеристик обрабатываемых данных (возможно, неточных), либо результатов таких их преобразований, что оказываются удобными для дальнейшего принятия решений. Следовательно, суть обработки данных – в выполняемых вычислениях.

Если исходные данные, подвергаемые обработке, неточны (например, являются результатами измерений), то неточны и любые результаты их обработки. Назначение данного курса заключается в том, чтобы научиться с помощью программных средств автоматически определять характеристики погрешности результатов вычислений с неточными данными, унаследованные ими от неточных исходных данных.

Программное обеспечение

Основной рекомендуемый инструмент для выполнения заданий – пакет математического моделирования Matlab, с которым Вы знакомы. При этом при желании можно воспользоваться любым другим инструментом программирования (C++, Python и др.). Дальнейшие примеры программирования относятся к Matlab.

Чтобы воспользоваться программным обеспечением Matlab в условиях дистанционного образовательного процесса, можно использовать один из трех взаимозаменяемых вариантов.

1. Получить студенческую лицензию на сайте разработчика.

https://ch.mathworks.com/campaigns/products/trials.html

2. Воспользоваться сервисами удаленного доступа СПбПУ, позволяющими получить доступ к университетскому серверу Matlab.

https://www.spbstu.ru/upload/it/VDI_for_users.pdf

3. Выполнять задания в свободно распространяемом пакете математического моделирования Octave, язык программирования которого в рамках предлагаемых заданий полностью идентичен языку программирования Matlab.

https://www.gnu.org/software/octave/

https://octave.sourceforge.io/

Для Octave есть online-версия для браузера:

https://octave-online.net/

Лабораторная работа № 5

Лабораторная работа посвящена сравнительному исследованию методов оценки центра обрабатываемой совокупности с точки зрения робастности получаемых оценок (устойчивости к наличию выбросов в данных).

Необходимо сравнить следующие статистические оценки центра совокупности:

– середина размаха,

– среднее арифметическое,

– усеченное среднее,

– медиана.

Работа включает: программирование аналитических оценок (см. лекции) наследственной погрешности для каждой оценки центра обрабатываемой совокупности, вызванной неточностью исходных данных (т.е. элементов совокупности); проведение статистического эксперимента методом Монте-Карло с целью получения стохастических оценок наследственной погрешности; сравнение аналитических оценок с численными оценками, полученными методом Монте-Карло; интерпретация результатов и выводы.

Постановка задачи

Пусть имеется совокупность значений $\mathbf{x} = (x_1,\ x_2,\ \dots,\ x_n)^{\mathrm{T}}$, полученных в результате выполнения измерений.

Пусть Δx_i – абсолютная погрешность величины x_i, т.е $\Delta x_i = x_i - x_{\text{ист}\,i}$, где $x_{\text{ист}\,i}$ – истинное значение измеряемой величины x_i. Пусть Δ_i – предельное (наибольшее возможное) значение модуля абсолютной погрешности величины x_i, то есть $|\Delta x_{\mathrm{i}}| < \Delta_i$.

Необходимо оценить наследственную погрешность результата оценки центра совокупности $\mathbf{x}$ указанными выше методами.

Пусть алгоритм обработки неточных данных задан функцией вида

$$y_{\text{ист}} = f(x_{\text{ист}\,1},\, x_{\text{ист}\,2},\ \dots,\, x_{\text{ист}\,n}),$$

где f – некоторая функция от элементов совокупности (в рамках лабораторной работы – вычисление центра совокупности действительных значений $\mathbf{x}_{\text{ист}}$), $y_{\text{ист}}$ – значения данной функции f.

Поскольку элементы совокупности $\mathbf{x}$ известны с погрешностью, то точно определить значение центра совокупности нельзя. Получаемая оценка неизбежно содержит наследственную погрешность Δy, обусловленную наличием погрешностей Δx_1, Δx_2, …, Δx_n. Кроме того, значение Δy будет зависеть от вида функции f, то есть от выбранного метода вычисления среднего.

Исходные данные

Принять в качестве исследуемой совокупности вектор значений из 99 случайных значений в диапазоне от (–1) до (+1), а последнее (сотое) значение – равным 1.589.

Принять следующие пределы относительных погрешностей: 1% – для всех значений, кроме последнего, (1, 2, …, 20)% – для последнего значения в указанной выше совокупности.

Таким образом, моделируется ситуация, когда одно значение в совокупности получено с существенно большей погрешностью, нежели остальные члены ряда, т.е. представляет собой статистический выброс.

Число итераций статистического эксперимента методов Монте-Карло задать порядка $N = 10^5 \div 10^6$.

Замечание: на этапе отладки программы полезно задать меньшее значение N, чтобы ускорить вычисления.

Задание к лабораторной работе № 5

Оценить значение центра для заданной совокупности $\mathbf{x}$: с помощью середины размаха, среднего арифметического, усеченного среднего, медианы.

Для каждой из оценок оценить аналитически и методом Монте-Карло наибольшее возможное значение Δ_y модуля отклонения Δy оценки значения центра совокупности **x** для заданных пределов погрешности исходных данных, приняв для усеченного среднего коэффициент $k = 0{,}2$.

Построить графики (ниже Δ_n – это предел абсолютной погрешности последнего значения в обрабатываемой совокупности, $n = 100$):

– зависимости Δ_y от Δ_n, где Δ_y оценивается по аналитическим формулам,

– зависимости Δ_y от Δ_n, где Δ_y оценивается методом Монте-Карло.

Повторить весь эксперимент и построить графики, введя в обрабатываемую совокупность дополнительно 1, 2, 3 выброса (их значения – на Ваше усмотрение). Повторить весь эксперимент и построить графики, изменив значение коэффициента k для оценки усеченным средним на значение, равное 0,1.

В разделе «Выводы» ответить на следующие вопросы (опираясь, в том числе, на результаты выполненного эксперимента).

а) Какая из оценок центра совокупности наименее робастная? Какая – наиболее робастная?

б) В чем причина схожей чувствительности к выбросам оценок среднего медианой и усеченным средним в настоящем примере?

в) Как увеличение и уменьшение значения k влияет на точность оценки центра совокупности усеченным средним? Как измерение значения k влияет на робастность данной оценки?

в) В каком случае оценка центра медианой предпочтительнее оценки усеченным средним?

Теоретические сведения, необходимые для выполнения работы

При выполнении лабораторной работы используйте следующие формулы для вычисления среднего по совокупности:

– середина размаха

$$x_{\text{sc}} = \frac{1}{2}\cdot\left(x_{(1)} + x_{(n)}\right),$$

– среднее арифметическое

$$\bar{x} = \frac{1}{n}\cdot\sum_{i=1}^{n} x_i,$$

– усеченное среднее

$$x_{\text{tr}} = \frac{1}{\left[(1-k)\cdot n\right]}\cdot\sum_{i=\left[0.5\cdot k\cdot n\right]}^{\left[(1-0.5\cdot k)\cdot n\right]+1} x_{(i)},$$

– медиана

$$x_{\text{med}} = x_{(0.5\cdot(n+1))}, \text{ если } n \text{ – нечетное},$$

$$x_{\text{med}} = \frac{1}{2}\cdot\left(x_{(n/2)} + x_{(n/2+1)}\right), \text{ если } n \text{ – четное},$$

где k – доля (%/100) элементов совокупности, используемых при оценке центра совокупности усеченным средним, $[x]$ – целая часть значения x, $x_{(i)}$ – i-ый элемент вариационного ряда, n – число элементов исследуемой совокупности.

Используйте следующие формулы для вычисления интервалов возможных значений погрешности оценки центра совокупнисти, унаследованной от исходных данных:

– для середины размаха:

$$l_1 = (x_1 - \Delta_1), l_2 = (x_2 - \Delta_2), \ldots, l_n = (x_n - \Delta_n),$$

$$r_1 = (x_1 + \Delta_1), r_2 = (x_2 + \Delta_2), \ldots, r_n = (x_n + \Delta_n),$$

$$x_{SC} \in \left[\frac{1}{2} \cdot \left(I_{(1)} + I_{(n)} \right), \frac{1}{2} \cdot \left(r_{(1)} + r_{(n)} \right) \right],$$

– для среднего арифметического:

$$\overline{x} \in \left[\overline{x} - \overline{\Delta}, \overline{x} + \overline{\Delta} \right],$$

– для усеченного среднего:

$$I_1 = \left(x_1 - \Delta_1 \right), I_2 = \left(x_2 - \Delta_2 \right), \ldots, I_n = \left(x_n - \Delta_n \right),$$

$$r_1 = \left(x_1 + \Delta_1 \right), r_2 = \left(x_2 + \Delta_2 \right), \ldots, r_n = \left(x_n + \Delta_n \right),$$

$$L_{x_{\mathrm{tr}}} = \frac{1}{\left[(1-k) \cdot n \right]} \cdot \sum_{i=\left[0.5 \cdot k \cdot n \right]}^{\left[(1-0.5 \cdot k) \cdot n \right]+1} I_{(i)},$$

$$R_{x_{\mathrm{tr}}} = \frac{1}{\left[(1-k) \cdot n \right]} \cdot \sum_{i=\left[0.5 \cdot k \cdot n \right]}^{\left[(1-0.5 \cdot k) \cdot n \right]+1} r_{(i)},$$

$$x_{\mathrm{tr}} \in \left[L_{x_{tr}}, R_{x_{tr}} \right],$$

– для медианы:

$$x_{\mathrm{med}} \in \left[L_{x_{\mathrm{med}}}, R_{x_{\mathrm{med}}} \right],$$

где, если n – нечетное,

$$L_{x_{\mathrm{med}}} = I_{\left(0.5 \cdot (n+1) \right)},$$

$$R_{x_{\mathrm{med}}} = r_{\left(0.5 \cdot (n+1) \right)},$$

а если n – четное, то

$$L_{x_{\mathrm{med}}} = \frac{1}{2} \cdot \left(I_{(n/2)} + I_{(n/2+1)} \right),$$

$$R_{x_{\mathrm{med}}} = \frac{1}{2} \cdot \left(r_{(n/2)} + r_{(n/2+1)} \right).$$

Вычислив границы интервалов, нетрудно найти предельное значение Δ_y модуля $|\Delta y|$ отклонения результата y оценки среднего по совокупности $\mathbf{x}$ для каждого из методов.

Среда реализации: любая на Ваш выбор.
По итогам реализации составьте краткий отчет, содержащий: титульный лист, текст задания, текст составленной Вами программы, полученные результаты, выводы.

В отчет прошу включить краткое описание задания, полученный результат, выводы.