Київський національний університет імені Тараса Шевченка
Кафедра теорії ймовірностей, статистики та актуарної математики

Р. Майборода

Комп'ютерна статистика. III курс

Рекомендації по виконанню індивідуальних робіт

Київ — 2016

Варіант	Множник	Приріст	Зернина
1	65539	0	2^{15}
2	65521	0	2^{10}
3	65527	0	2^{7}
4	75831	0	2^{15}
5	65547	6000	2^{10}
6	65549	6000	2^{7}
7	65539	6000	2^{15}
8	65521	3000	2^{10}
9	65527	3000	2^{7}
10	75831	3000	2^{5}

Таблиця 1: Параметри для генерації рівномірних послідовностей

Робота 1
ГЕНЕРАЦІЯ ПСЕВДОВИПАДКОВИХ ЧИСЕЛ

1. Розробити функцію `R` , що реалізує лінійний конгруентний генератор рівномірної псевдовипадкової послідовності з цілочисловою арифметикою. Функція без параметрів при кожному виклику повинна видавати одне число — черговий елемент псевдовипадкової послідовності. Генератор оформити у двох варіантах:

(а) З параметрами, що відповідають генератору Парка та Мілера.

(б) з модулем $m = 2^{31}$ та іншими параметрами, обраними з таблиці 1 за номером варіанту.

2. Використовуючи функції, розроблені у п.1, згенерувати послідовності з 500 елементів обома генераторами і перевірити їх якість із застосуванням:

(а) порівняння емпіричної функції розподілу з теоретичною;

(б) діаграми послідовності;

(в) діаграм пар та трійок елементів послідовності.

Зробити висновок про можливість застосування отриманих генераторів.

3. Використовуючи генератор, обраний у п.2, розробити функцію `R` , що генерує псевдовипадкові послідовності із розподілом, вказаним для вашого варіанту у таблиці 2. (У цій таблиці ξ та η — незалежні випадкові

Ваіант	Розподіл
1	Суміш Exp(0.5), Exp(1) з ймовірністю змішування $p = 0.4$
2	Суміш N(1,1), N(-1,1) з ймовірністю змішування $p = 0.6$
3	Суміш N(1,1), N(1,3) з ймовірністю змішування $p = 0.6$
4	Логнормальний розподіл з $m = 1$, $\sigma^2 = 0.75$
5	Симетричний трикутний розподіл на [-1.1]
6	$\xi + \eta$, $\xi \sim Exp(0.5)$, $\eta \sim Unif[-1, 1]$
7	$\xi + \eta$, $\xi \sim N(0, 1)$, $\eta \sim Unif[-1, 1]$
8	Розподіл Коші (стандартний)
9	$\xi + \eta$, $\xi \sim Exp(1)$, $\eta \sim Unif[0, 1]$
10	Суміш N(1,1), N(-1,0.5) з ймовірністю змішування $p = 0.5$

Таблиця 2: Розподіли для генерації псевдовипадкових послідовностей

величини).

4. Перевірити якість роботи генератора з п. 3, порівнявши емпіричну та теоретичну функції розподілу.

5. Згенерувати послідовність з тим же розподілом, використовуючи стандартні функції генерації псевдовипадкових чисел з `R` . Порівняти якість отриманої послідовності з послідовністю п. 4.

Додаткове завдання: Розробити функцію, що генерує рівномірну послідовність методом Фібоначчі. Перевірити її якість як описано у п.2. Використати отриману послідовність для генерації послідовності із заданим розподілом як у п.3. Порівняти результат з тим, що було згенеровано лінійним конгруентним генератором.

Література

Основна:

1. Р. Майборода Комп'ютерна статистика. Розділ "Генерація псевдовипадкових послідовностей".

Додаткова:

2. William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P.Flannery Numerical Recipes in C: The Art of Scientific Computing (1992) Cambridge University Press New York, NY, USA

3. Кнут Д. Э., Искусство программирования. Том 2. Получисленные методы — Вильямс. 2001.

4. William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P.Flannery Numerical Recipes in C: The Art of Scientific Computing (1992)

Робота 2
ВІЗУАЛЬНИЙ РОЗПОДІЛУ АНАЛІЗ ДАНИХ

1. Прочитайте у `R` дані з файлу `distr`N`.txt`, де N — номер Вашого варіанту. Файл можна взяти з архіву за адресою:

`http://probability.univ.kiev.ua/userfiles/mre/compsta3.rar`

2. Побудуйте гістограму прочитаних даних разом із щільністю розподілу а також QQ і PP-діаграми, використовуючи для порівняння такі розподіли:

(а) нормальний;

(б) логнормальний;

(в) експоненційний;

(г) χ^2.

Параметри розподілів спробуйте підібрати на око, використовуючи будь-які евристичні міркування.

Виберіть теоретичний розподіл, який, на Вашу думку, найкраще описує розподіл даних.

3. Побудуйте з обраним теоретичним розподілом QQ-діаграму з прогнозними інтервалами. (Використайте рівень значущості $\alpha = 0.05$). Зробіть остаточний висновок про те, чи відповідає розподіл даних теоретичному.

4. (Додаткове завдання) Напишіть функцію `R` , що будує гістограму відносних частот з графіком щільності обраного Вами теоретичного розподілу і відповідними прогнозними інтервалами для відносних частот. Перевірте роботу цієї функції на модельованих даних та на даних з Вашого варіанту.

Література

Основна:

1. Р. Майборода Комп'ютерна статистика. Розділ “Методи графічного аналізу одновимірних даних”.

Додаткова:

2. М.В.Карташов Ймовірність, процеси, статистика.

3. W. N. Venables and B. D. Ripley Modern Applied Statistics with S. Forth edition. Springer, 2002.

Робота 3
ОЦІНЮВАННЯ МЕТОДОМ МОМЕНТІВ

1. Розробити оцінку методу моментів $\hat{\vartheta}^{MM}$ для оцінювання невідомого параметру ϑ за кратною вибіркою з розподілом що відповідає номеру варіанту:

1. Суміш двох експоненційних розподілів з інтенсивностями λ_1 та λ_2 і ймовірністю змішування p. Невідомий параметр $\vartheta = \lambda_1$. Значення відомих параметрів — $p = 0.4$, $\lambda_2 = 0.5$.

2. Суміш двох експоненційних розподілів з інтенсивностями λ_1 та λ_2 і ймовірністю змішування p. Невідомий параметр $\vartheta = p$. Значення відомих параметрів — $\lambda_1 = 1$, $\lambda_2 = 0.5$.

3. суміш двох нормальних розподілів з математичними сподіваннями μ_1, μ_2, середньоквадратичними відхиленнями σ_1, σ_2 та ймовірністю змішування p. Невідомий параметр $\vartheta = \mu_1$. Значення відомих параметрів — $\sigma_1 = 1$, $\mu_2 = 1$, $\sigma_2 = 0.75$, $p = 0.6$.

4. суміш двох нормальних розподілів з математичними сподіваннями μ_1, μ_2, середньоквадратичними відхиленнями σ_1, σ_2 та ймовірністю змішування p. Невідомий параметр $\vartheta = \sigma_1$. Значення відомих параметрів — $\mu_1 = 1$, $m_2 = 0$, $\sigma_2 = 0.75$, $p = 0.6$.

5. суміш двох нормальних розподілів з математичними сподіваннями μ_1, μ_2, середньоквадратичними відхиленнями σ_1, σ_2 та ймовірністю змішування p. Невідомий параметр $\vartheta = p$. Значення відомих параметрів — $\mu_1 = 1$, $m_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 0.75$, p.

6. $\xi + \eta$, де ξ і η незалежні. ξ має нормальний розподіл з середнім μ та середньоквадратичним відхиленням σ. η має рівномірний розподіл на [-1,1]. Невідомий параметр $\vartheta = \mu$. Значення відомого параметру — $\sigma = 1$.

7. $\xi + \eta$, де ξ і η незалежні. ξ має нормальний розподіл з середнім μ та середньоквадратичним відхиленням σ. η має рівномірний розподіл на [-1,1]. Невідомий параметр $\vartheta = \sigma$. Значення відомого параметру — $\mu = 1$.

8. $\xi+\eta$, де ξ і η незалежні. ξ має експоненційний розподіл з інтенсивністю λ . η має рівномірний розподіл на [-1,1]. Невідомий параметр $\vartheta = \lambda$.

9. Трикутний розподіл на інтервалі [0,2] з модою a. Невідомий параметр $\vartheta = a$.

10. Логнормальний розподіл з параметрами μ і σ. Невідомий параметр $\vartheta = \mu$. Значення відомого параметру $\sigma = 1$.

2. Розробити функцію `R` з назвою `EstMM`, що реалізує оцінку $\hat{\vartheta}^{MM}$. Аргумент функції — вибірка, значення функції — оцінка за цією вибіркою. Перевірити роботу функції на модельованій вибірці.

3. Підрахувати коефіцієнт розсіювання оцінки за теоретичною формулою.

4. Згенерувати $B = 1000$ вибірок обсягу n із заданого розподілу. По кожній вибірці оцінити невідомий параметр і записати у масив оцінок. За масивом оцінок оцінити змішення та коефіцієнт розсіювання. Порівняти з теоретичним значенням коефіцієнту розсіювання. Побудувати гістограму оцінок з відповідною нормованою щільністю нормального розподілу.

Обсяг вибірки n обирати рівним 50, 100, 500 і т.д., намагаючись отримати хорошу відповідність емпіричних результатів асимптотичних значень.

Значення невідомого параметру для моделювання обрати самостійно. По можливості, провести експерименти з різними значеннями.

5. Розробити функцію `R` що підраховує асимптотичний довірчий інтервал для невідомого параметру за допомогою коефіцієнту розсіювання.

6. Перевірити реальний рівень значущості отриманого довірчого інтервалу при номінальному рівні значущості $\alpha = 0.05$ використовуючи модельовані вибірки з п. 4.

Додаткове завдання: Реалізувати оцінку методу медіан (квантилів) для невідомого параметру. Порівняти її якість з якістю оцінки методу моментів.

Література

Основна:

1. Р. Майборода Комп'ютерна статистика. Підрозділи "Оцінки узагальненого методу моментів", "Асимптотична нормальність і матриця розсіювання оцінок", "Довірчі інтервали".

Додаткова:
2. М.В.Карташов Ймовірність, процеси, статистика.
3. Боровков А.А. Математическая статистика. - Наука, Москва, 1984.

Робота 4
ОЦІНКИ МЕТОДУ НАЙБІЛЬШОЇ ВІРОГІДНОСТІ

1. Розробити оцінку методу найбільшої вірогідності $\hat{\vartheta}^{MLE}$ для оцінювання невідомого параметру ϑ за кратною вибіркою з розподілом який був визначений у роботі 3.

2. Оцінку реалізувати у вигляді функції `R` з назвою `EstMLE`. Аргумент функції — вибірка, значення функції — оцінка $\hat{\vartheta}^{MLE}$. Для знаходження максимуму функції вірогідності при потребі використати метод Ньютона. На роль початкового наближенння взяти оцінку методу моментів з роботи 3. Перевірити роботу функції на модельованій вибірці.

3. Підрахувати інформацію за Фішером та коефіцієнт розсіювання оцінки методу найбільшої вірогідності. Знайти відносну асимптотичну ефективність оцінок методу моментів та методу найбільшої вірогідності.

4. Згенерувати $B = 1000$ вибірок обсягу n із заданого розподілу. По кожній вибірці оцінити невідомий параметр і записати у масив оцінок. За масивом оцінок оцінити змішення та коефіцієнт розсіювання. Порівняти з теоретичним значенням коефіцієнту розсіювання. Побудувати гістограму оцінок з відповідною нормованою щільністю нормального розподілу.

5. Порівняти оцінки методу моментів з роботи 3 та методу найбільшої вірогідності. Зробити висновок про доцільність практичного використання цих оцінок.

Література

Основна:

1. Р. Майборода Комп'ютерна статистика. Підрозділи “Оцінки методу найбільшої вірогідності”, “Асимптотична нормальність і матриця розсіювання оцінок”.

Додаткова:

2. М.В.Карташов Ймовірність, процеси, статистика.

3. Боровков А.А. Математическая статистика. - Наука, Москва, 1984.

Робота 5
ПЕРЕВІРКА ГІПОТЕЗ

1. Розробити функцію `R` що підраховує за вибіркою відношення вірогідності для перевірки простої гіпотези проти простої альтернативи за ймовірнісною моделлю, що обирається за номером варіанту:

1. H_0 : Exp(7), H_1 : Exp(9), n=80
2. H_0 : Beta(2,5), H_1 : Beta(x,2,6), n=65
3. H_0 : Poisson(0.5), H_1 : Poisson(0.3), n=75
4. H_0 : Binom(0.5,2), H_1 : Binom(0.6,2), n=65
5. H_0 : Beta(4,3), H_1 Beta(5,3) : n=50
6. H_0 : Poisson(0.3), H_1 : Poisson(0.5), n=45
7. H_0 : Exp(6), H_1 : Exp(5), n=55
8. H_0 : Beta(2,5), H_1 : Beta(2,4), n=50
9. H_0 : Binom(0.7,2), H_1 : Binom(0.5,2), n=25
10. H_0 : Poisson(3), H_1 : Poisson(4), n=45

(тут Exp(λ) – експоненційний розподіл з інтенсивністю λ; Beta(α, β) – бета-розподіл з параметрами α, β; Poisson(λ) – Пуассонів розподіл з параметром λ; Binom(p,m) – біноміальний розподіл з ймовірністю успіху p та кількістю випробувань m, n — обсяг вибірки).

2. Методом імітаційного моделювання визначити поріг тесту відношення вірогідності для перевірки H_0 проти H_1 з рівнем значущості $\alpha = 0.05$ при заданому обсязі вибірки n.

3. Методом імітаційного моделювання визначити ймовірність помилки другого роду для розробленого тесту.

4. Підібрати розмір найменший вибірки n, при якому тест матиме ймовірності помилок першого і другого роду, що не перевищують 0.05.

Література

Основна:

1. Р. Майборода Комп'ютерна статистика. Розділ "Перевірка статистичних гіпотез".

Додаткова:
2. М.В.Карташов Ймовірність, процеси, статистика.
3. Боровков А.А. Математическая статистика. - Наука, Москва, 1984.