

О. А. Сердюк, В. М. Трояновский (Москва, Зеленоград, М-ГИЭТ (ТУ)). **Статистические проблемы оценивания экспериментальных данных с помощью гистограмм.**

Статистические методы широко используются при обработке экспериментальных данных. Они опираются на глубокие теоретические исследования с привлечением теории множеств. Часто мощность множеств предполагается бесконечной. Вместе с этим, экспериментаторы работают с реализациями ограниченной длины, и поэтому получаемые оценки всегда имеют статистические флуктуации. Для оценки таких флуктуаций в классической статистике разработан мощный аппарат проверки гипотез, однако, результаты даются лишь в виде вероятностных оценок, и на экспериментатора перекладывается ответственность принятия окончательного решения даже в том случае, когда вероятность подтверждения гипотезы оказывается не слишком высокой.

Возникает задача получения количественных характеристик точности оценок, рассчитываемых на основе экспериментальных данных. Сложность задачи усугубляется тем, что характеристики и зависимости, полученные из эксперимента, как правило, выглядят существенно более «изрезанными», чем кривые, рассчитанные на основе теории множеств. Для примера укажем, что при сложении 2-х, 3-х, ..., 5-ти случайных величин с исходным равномерным распределением на интервале $[0,1]$ закон распределения для их суммы очень быстро приближается к нормальному, в то время как расчет гистограмм (как для равномерного, так и для нормального распределения) дает удовлетворительное приближение к теоретическим кривым лишь при объемах данных, в несколько тысяч отсчетов [1]. Это является одним из проявлений различий в усреднении по множеству и вдоль единственной реализации [2].

Теоретический анализ указанных различий применительно к использованию гистограмм является достаточно сложной задачей. Однако вычислительный эксперимент позволяет получить желаемые оценки разброса ординат гистограммы с точностью, приемлемой для инженерных применений.

С этой целью была разработана программа построения гистограмм по данным, получаемым от датчика случайных чисел. Фиксировался разброс величин относительно теоретического значения в каждом из 10 интервалов в гистограмме. (Для простоты в качестве начального приближения использовалось нормальное распределение). Для набора статистики эксперимент повторялся 30 раз, после чего фиксировались минимальное и максимальное значения, затем производилась новая серия экспериментов с иным объемом данных. Обработка полученных результатов показала, что имеет место достаточно типичная для статистических методов сходимость результатов с ростом числа отсчетов. Это позволило предложить эмпирическую формулу для оценки границ пребывания ординат нормированной гистограммы в виде:

$$U_{\max} - U_{\min} \approx 0,7/\sqrt{m},$$

где m — объем выборки.

В докладе приводятся подробные сведения о методике и результатах эксперимента.

СПИСОК ЛИТЕРАТУРЫ

1. *Трояновский В. М.* Информационно-управляющие системы и прикладная теория случайных процессов: Учебное пособие. М.: Гелиос АРВ, 2004, 304 с.
2. *Вентцель Е. С., Овчаров Л. А.* Теория случайных процессов и ее инженерные приложения. М.: Наука, 1991, 248 с.
3. *Трояновский В. М.* Особенности усреднения по множеству и по времени в задачах анализа и синтеза информационно-управляющих систем. — Обзорение прикл. и промышл. матем., 2007, т. 14, в. 3, с. 567–568.