ОБРАБОТКА И АНАЛИЗ РЕЗУЛЬТАТОВ МОДЕЛИРОВАНИЯ

Особенности статистической обработки результатов ЭВМ

При выборе методов обработки существенную роль играют три особенности машинного эксперимента с моделью системы *S*.

1. Возможность получать при моделировании системы *S* на ЭВМ большие выборки позволяет количественно оценить характеристики процесса функционирования системы, но превращает в серьезную проблему хранение промежуточных результатов моделирования. Эту проблему можно решить, используя рекуррентные алгоритмы обработки, когда оценки вычисляют по ходу моделирования.

2. Сложность исследуемой системы *S* при ее моделировании на ЭВМ часто приводит к тому, что априорное суждение о характеристиках процесса функционирования системы, например о типе ожидаемого распределения выходных переменных, является невозможным. Поэтому при моделировании систем широко используются непараметрические оценки и оценки моментов распределения.

3. Блочность конструкции машинной модели $M_м$ и раздельное исследование блоков связаны с программной имитацией входных переменных для одной частичной модели по оценкам выходных переменных, полученных на другой частичной модели. Если ЭВМ, используемая для моделирования, не позволяет воспользоваться переменными, записанными на внешние

носители, то следует представить эти переменные в форме, удобной для построения алгоритма их имитации.

При исследовании сложных систем и большом числе реализаций N в результате моделирования на ЭВМ получается значительный объем информации о состояниях процесса функционирования системы. Поэтому необходимо так организовать в процессе вычислений фиксацию и обработку результатов моделирования, чтобы оценки для искомых характеристик формировались постепенно по ходу моделирования, т. е. без специального запоминания всей информации о состояниях процесса функционирования системы S.

Если при моделировании процесса функционирования конкретной системы S учитываются случайные факторы, то и среди результатов моделирования присутствуют случайные величины. В качестве оценок для искомых характеристик рассчитывают средние значения, дисперсии, корреляционные моменты и т. д.

При обработке результатов моделирования можно подойти к оценке вероятностей возможных значений случайной величины, т. е. закона распределения. Область возможных значений случайной величины η разбивается на n интервалов. Затем накапливается количество попаданий случайной величины в эти интервалы m_k, $k=1, n$. Оценкой для вероятности попадания случайной величины в интервал с номером k служит величина m_k/N. Таким образом, при этом достаточно фиксировать n значений m_k при обработке результатов моделирования на ЭВМ.

Для оценки среднего значения случайной величины η накапливается сумма возможных значений случайной величины *y_k, k=1, N,* которые она принимает при различных реализациях. Тогда среднее значение

$$\bar{y} = (1/N)\sum_{k=1}^{N} y_k .$$

При этом ввиду несмещенности и состоятельности оценки

$$M[\bar{y}] = M[\eta] = \mu_\eta ;$$

$$D[\bar{y}] = D[\eta]/N = \sigma_\eta^2 / N .$$

В качестве оценки дисперсии случайной величины η при обработке результатов моделирования можно использовать

$$S_b^2 = \sum_{k=1}^{N} (y_k - \bar{y})^2 / N$$

При обработке результатов машинного эксперимента с моделью *$M_м$* наиболее часто возникают следующие задачи: определение эмпирического закона распределения случайной величины, проверка однородности распределений, сравнение средних значений и дисперсий переменных, полученных в результате моделирования, и т. д. Эти задачи с точки зрения математической статистики являются типовыми задачами по проверке статистических гипотез.

Задача определения эмпирического закона распределения случайной величины наиболее общая из перечисленных, но для правильного решения требует большого числа реализаций *N.* В этом случае по результатам машинного эксперимента находят

значения выборочного закона распределения *$F_э(y)$* (или функции плотности *$f_э(y)$*) и выдвигают нулевую гипотезу *H_0*, что полученное эмпирическое распределение согласуется с каким-либо теоретическим распределением. Проверяют эту гипотезу *H_0* с помощью статистических критериев согласия Колмогорова, Пирсона, Смирнова и т. д., причем необходимую в этом случае статистическую обработку результатов ведут по возможности в процессе моделирования системы *S* на ЭВМ.

Критерий согласия Колмогорова основан на выборе в качестве меры расхождения *U* величины $D = max\left[F_э(y) - F(y)\right]$.

Из теоремы Колмогорова следует, что $\delta = D\sqrt{N}$ **при** $N \rightarrow \infty$ **имеет функцию распределения**

$$F(z) = P\{\delta < z\} = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2z^2}, z > 0.$$

Если вычисленное на основе экспериментальных данных значение 5 меньше, чем табличное значение при выбранном уровне значимости у, то гипотезу *Я*$_0$ принимают, в противном случае расхождение между *$F_э(y)$* и *$F(y)$* считается неслучайным гипотеза *H_0* отвергается.

Критерий Колмогорова для обработки результатов моделирования целесообразно применять в тех случаях, когда известны все параметры теоретической функции распределения. Недостаток использования этого критерия связан с необходимостью фиксации в памяти ЭВМ для определения *D* всех статистических частот с целью их упорядочения в порядке возрастания.

Критерий согласия Пирсона основан на определении в качестве меры расхождения *U* величины

$$\chi^2 = \sum_{i=1}^{d} \left(m_i - N_{pi}\right) / \left(N_{pi}\right),$$

где m_i — количество значений случайной величины η, попавших в *i*-й подынтервал; p_i — вероятность попадания случайной величины η в *i*-й подынтервал, вычисленная из теоретического распределения; *d* — количество подынтервалов, на которые разбивается интервал измерения в машинном эксперименте.

При $N \to \infty$ закон распределения величины *U*, являющейся мерой расхождения, зависит только от числа подынтервалов и приближается к закону распределения χ^2 (хи-квадрат) с *(d-r-1)* степенями свободы, где *r* — число параметров теоретического закона распределения.

Из теоремы Пирсона следует, что, какова бы ни была функция распределения *F(y)* случайной величины η, при $N \to \infty$ распределение величины χ^2 имеет вид

$$F_k(z) = P\left\{\chi^2 < z\right\} = 1 / \left[2^{k/2}\, \Gamma(k/2)\right] \int_0^z e^{-t/2} t^{(k/2-1)} dt,\ z > 0$$

где *Γ(k/2)* — гамма-функция; z — значение случайной величины χ^2, *k = d-r-1* — число степеней свободы. Функции распределения $F_k(z)$ табулированы.

По вычисленному значению $U = \chi^2$ и числу степеней свободы *k* с помощью таблиц находится вероятность $P\left\{\chi_\tau^2 \geq \chi^2\right\}$.Если эта вероятность превышает некоторый уровень значимости γ, то

считается, что гипотеза H_0 о виде распределения не опровергается результатами машинного эксперимента.

Для принятия или опровержения гипотезы выбирают некоторую случайную величину *U,* характеризующую степень расхождения теоретического и эмпирического распределения, связанную с недостаточностью статистического материала и другими случайными причинами. Закон распределения этой случайной величины зависит от закона распределения случайной величины η и числа реализаций *N* при статистическом моделировании системы *S.* Если вероятность расхождения теоретического и эмпирического распределений $P\{U_T \geq U\}$ велика в понятиях применяемого критерия согласия, то проверяемая гипотеза о виде распределения H_0 не опровергается. Выбор вида теоретического распределения *F(y)* (или *f(y)*) проводится по графикам (гистограммам) $F_э(y)$ (или $f_э(y)$), выведенным на печать или на экран дисплея.

Хотя рассмотренные оценки искомых характеристик процесса функционирования системы *S,* полученные в результате машинного эксперимента с моделью M_M, являются простейшими, но охватывают большинство случаев, встречающихся в практике обработки результатов моделирования системы для целей ее исследования и проектирования.

Корреляционный анализ результатов моделирования

С помощью *корреляционного анализа* исследователь может установить, насколько тесна связь между двумя (или более) случайными величинами, наблюдаемыми и фиксируемыми при моделировании конкретной системы *S.* Корреляционный анализ

результатов моделирования сводится к оценке разброса значений η относительно среднего значения $\bar{y}$, т. е. к оценке силы корреляционной связи. Существование этих связей и их тесноту можно для схемы корреляционного анализа $y = M[\eta/\xi = x]$ выразить при наличии линейной связи между исследуемыми величинами и нормальности их совместного распределения с помощью коэффициента корреляции.

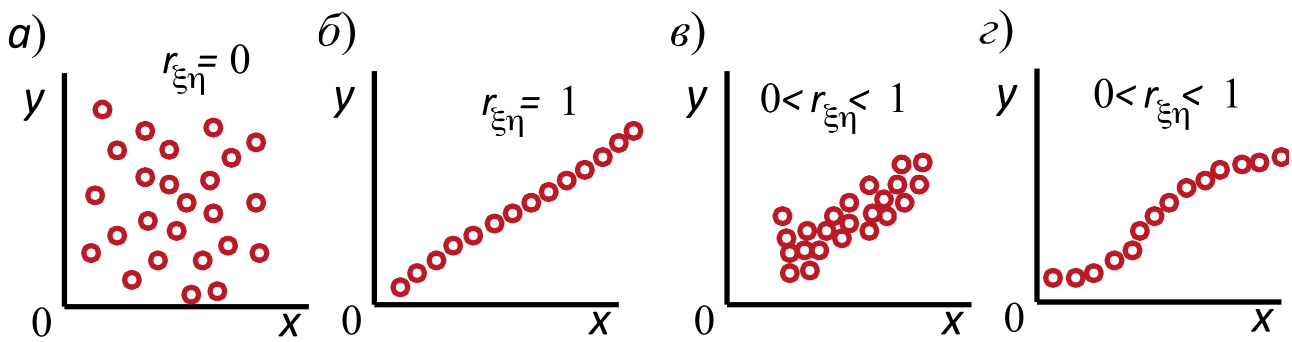

Рис.1. Различные случаи корреляции переменных

Для того чтобы оценить точность полученной при обработке результатов моделирования системы *S* оценки $r_{\xi\eta}$, целесообразно ввести в рассмотрение коэффициент

$$w = \ln [(1+ r_{\xi\eta})/(1-r_{\xi\eta})]/2,$$

причем w приближенно подчиняется гауссовскому распределению со средним значением и дисперсией:

$$\mu_w = ln\left[\left(1 + r_{\xi\eta}\right)/\left(1 - r_{\xi\eta}\right)\right]/2$$

$$\sigma_w^2 = 1/\left(N - 3\right)$$

Из-за влияния числа реализаций при моделировании *N* на оценку коэффициента корреляции необходимо убедиться в том, что $0 \le r_{\xi\eta} \le 1$ действительно отражает наличие статистически значимой корреляционной зависимости между исследуемыми

переменными модели *M_M*. Это можно сделать проверкой гипотезы *H_0*: *$r_{\xi\eta}$*=0. Если гипотеза *H_0* при анализе отвергается, то корреляционную зависимость признают статистически значимой. Очевидно, что выборочное распределение введенного в рассмотрение коэффициента *w* при *$r_{\xi\eta}$*= 0 является гауссовским с нулевым средним μ_w = 0 и дисперсией $\sigma_w^2 = (N-3)^{-1}$.

При анализе результатов моделирования системы *S* важно отметить то обстоятельство, что даже если удалось установить тесную зависимость между двумя переменными, то отсюда еще непосредственно не следует их причинно-следственная взаимообусловленность. Возможна ситуация, когда случайные ξ и η стохастически зависимы, хотя причинно они являются для системы *S* независимыми. При статистическом моделировании наличие такой зависимости может иметь место, например, из-за коррелированности последовательностей псевдослучайных чисел, используемых для имитации событий, положенных в основу вычисления значений *x* и *y*.

Таким образом, корреляционный анализ устанавливает связь между исследуемыми случайными переменными машинной модели и оценивает тесноту этой связи. Однако в дополнение к этому желательно располагать моделью зависимости, полученной после обработки результатов моделирования.

Регрессионный анализ результатов моделирования.

Регрессионный анализ **дает возможность построить модель, наилучшим образом соответствующую набору данных, полученных в ходе машинного эксперимента с системой *S*. Под наилучшим соответствием понимается минимизированная**

функция ошибки, являющаяся разностью между прогнозируемой моделью и данными эксперимента. Такой функцией ошибки при регрессионном анализе служит сумма квадратов ошибок.

Дисперсионный анализ результатов моделирования.

При обработке и анализе результатов моделирования часто возникает задача сравнения средних выборок. Если в результате такой проверки окажется, что математическое ожидание совокупностей случайных переменных $\{y^{(1)}\}$, $\{y^{(2)}\}$, …, $\{y^{(n)}\}$ отличается незначительно, то статистический материал, полученный в результате моделирования, можно считать однородным (в случае равенства двух первых моментов). Это дает возможность объединить все совокупности в одну и позволяет существенно увеличить информацию о свойствах исследуемой модели M_M, а следовательно, и системы *S.* Попарное использование для этих целей критериев Смирнова и Стьюдента для проверки нулевой гипотезы затруднено в связи с наличием большого числа выборок при моделировании системы. Поэтому для этой цели используется *дисперсионный анализ*.

Дисперсионный анализ позволяет вместо проверки нулевой гипотезы о равенстве средних значений выборок проводить при обработке результатов моделирования проверку нулевой гипотезы о тождественности выборочной и генеральной дисперсий.

Возможны и другие подходы к анализу и интерпретации результатов моделирования, но при этом необходимо помнить, что их эффективность существенно зависит от вида и свойств конкретной моделируемой системы *S.*