УДК 651.5.015

НЕКОТОРЫЕ ВОПРОСЫ ОЦЕНКИ АДЕКВАТНОСТИ МОДЕЛИ

Е. П. Черногоров

Рассматриваются вопросы оценки адекватности математической модели с помощью проверки статистических гипотез о принадлежности одной генеральной совокупности реализаций эксперимента и математического моделирования этого эксперимента. Для оценки статистических характеристик модели предлагается метод имитационного моделирования.

Ключевые слова: модель; математическое ожидание; дисперсия; имитационное моделирование.

При математическом моделировании сложных технических систем с. понятие адекватности модели может служить оценкой качества модели и ее пригодности лишь для тех целей, для которых предназначена модель. Кроме того, источник информации о действительности подвержен воздействию шумов в измерительной системе, поэтому параметры и поведение системы нам не известны точно.

При определении степени адекватности математической модели мы должны иметь в распоряжении результат реакции реальной системы на некоторое входное воздействие, а затем сравнить эту реакцию с результатом математического моделирования при том же входном воздействии.

Можно считать, что модель системы позволяет нам моделировать некоторую генеральную совокупность. Имея отдельный эксперимент, воспроизводящий «жизнь» некоторой системы, мы можем утверждать, что имеем отдельную реализацию этой генеральной совокупности. На модели, воспроизводя условия эксперимента, мы также получаем реализацию генеральной совокупности. Количественная оценка адекватности будет заключаться в проверке статистических гипотез о принадлежности этих реализаций одной и той же генеральной совокупности.

Итак, имеем единичные реализации некоторых случайных процессов. Показания прибора, регистрирующего результаты эксперимента, можно представить в виде:

$$P(t) = \bar{P}(t) + \delta P(t),$$

где $\bar{P}(t)$ – истинное значение параметра, нам неизвестное; $\delta P(t)$ – помеха, вызванная шумом в измерительном тракте. Пусть эта помеха представляет собой некоторый стационарный эргодический случайный нормальный процесс с нулевым матожиданием и дисперсией σ^2.

$$\delta P(t) \sim N(0, \sigma).$$

σ^2 легко оценить по результатам тарировки измерительного тракта.

Модель, описывающая изменение параметра в ходе эксперимента, выдает результат в виде временного ряда $P_{\text{М}}(t)$ причем:

$$P_{\text{М}}(t) = \bar{P}_{\text{М}}(t) + \delta P_{\text{М}}(t),$$

где $\bar{P}_{\text{М}}(t)$ – матожидание результата, которое было бы на выходе модели, если бы мы точно знали входное воздействие на систему в ходе эксперимента; $\delta P_{\text{М}}(t)$ – ошибка модели, связанная с неточным знанием входных воздействий на систему. Предположим, что ошибка эта некоторый стационарный эргодический нормальный процесс с нулевым матожиданием и дисперсией $\sigma_{\text{М}}^2$:

$$\delta P_{\text{М}}(t) \sim N(0, \sigma_{\text{М}}),$$

Для подсчета дисперсии ошибки моделирования можно воспользоваться зависимостью:

$$\sigma_{\text{М}}^2 = \sum_i \left(\frac{\partial P_{\text{М}}}{\partial \beta_i} \right)^2 \sigma_{\beta i}^2,$$

где β_i – параметр или вход модели, известный нам с погрешностью, дисперсия которой $\sigma_{\beta\text{i}}^2$. Предполагается, что β_i не коррелированы между собой.

Теперь для разности $P(t) - P_{\text{М}}(t)$ можем записать:

$$P(t) - P_{\text{М}}(t) = \bar{P}(t) - \bar{P}_{\text{М}}(t) + \delta P_{\Sigma}(t),$$

где $\delta P_{\Sigma}(t)$ – некоторая эргодическая стационарная случайная функция с нулевым матожиданием и дисперсией:

$$\sigma_{\Sigma}^2 = \sigma_{\text{М}}^2 + \sigma^2,$$

$$\delta P_{\Sigma}(t) \sim N(0, \sigma_{\Sigma}).$$

Выдвигается нулевая гипотеза H_0: $\bar{P}(t) = \bar{P}_{\text{М}}(t)$.

В этом случае $\Delta P(t) = P(t) - P_{\text{М}}(t)$ – некоторый нормальный стационарный эргодический процесс, параметры которого должны совпадать с параметрами процесса $\delta P_{\Sigma}(t)$:

$$\Delta P(t) \sim N\Big(M\big[\Delta P(t)\big], \sigma\big[\Delta P(t)\big]\Big).$$

Задача состоит в том, чтобы проверить случайность или неслучайность расхождения $\Delta P(t)$ и $\delta P_{\Sigma}(t)$ Проверка нулевой гипотезы при наличии реализации $\Delta P(t)$ будет состоять из следующих операций:

1. Проверка гипотезы об эргодичности процесса $\Delta P(t)$, Как известно, достаточным условием эргодичности стационарной случайной функции является стремление к нулю ее автокорреляционной функции при $\tau \to \infty$:

$$k_x(\tau) = \frac{1}{n-m} \sum_{i=1}^{n+m} \left(x\left(t_i - m_x\right)\right)\left(x\left(t_{i+m} - m_x\right)\right),$$

где $m_x = \dfrac{\sum_{i=1}^{n} x(t_i)}{m}$; $x(t_i) = \Delta P(t_i)$ – значение исследуемой функции.

2. После проверки гипотезы об эргодичности, при эргодичности процесса $\Delta P(t)$, можно определить оценки этого процесса по одной его реализации. Оценку матожидания найдем как среднее по времени:

$$m_i = M\left(\Delta P(t_i)\right) = \frac{1}{n} \sum_{i=1}^{n} \Delta P(t_i).$$

Оценку дисперсии можно определить как:

$$S_1^2 = S^2\left[\Delta P(t_i)\right] = \frac{1}{n} \sum_{i=1}^{n} \left(\Delta P(t_i) - m_i\right)^2.$$

Для процесса $\delta P_{\Sigma}(t)$ имеем матожидание:

$$m_2 = M\left[\delta P_{\Sigma}(t)\right] = 0$$

и дисперсию:

$$S_2^2 = S_2^2\left[\delta P_{\Sigma}(t_i)\right] = \frac{1}{n} \sum_{i=1}^{n} \sigma_{\Sigma}^2(t_i).$$

При больших n ($n > 30$) эти оценки можно считать несмещенными, при $n < 30$ вместо n следует брать $\nu = n - 1$ – число степеней свободы.

Проверка гипотезы H_0 разбивается на 2 этапа.

1. Проверка гипотезы о равенстве дисперсий процессов $\Delta P(t)$ и $\delta P_{\Sigma}(t)$:

$$H_0': \quad \sigma^2\left[\Delta P(t)\right] = \sigma^2\left[\delta P(t)\right].$$

Для сравнения дисперсий применяют критерий Фишера (дисперсионное отношение – отношение оценок дисперсий S_1^2 и S_2^2).

Проверяется выполнение условия:

$$F_{\left(1-\frac{\alpha}{2}\right)}(n) \geq \frac{S_1^2}{S_2^2}. \tag{1}$$

В числителе берется большая из оценок S_1^2 и S_2^2. В случае выполнения условия (1) принимается гипотеза H_0'; в случае невыполнения – альтернативная гипотеза.

На следующем этапе проверяется основная гипотеза H_0'': $m_1 = m_2 = 0$.

Здесь возможны два направления,

1. $\sigma^2\left[\Delta P\right] = \sigma^2\left[\delta P_\Sigma\right]$.

В этом случае m_1 подчиняется нормальному закону распределения, причем параметры этого распределения следующие:

$$N = \left(0, S^*\left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right)^{\frac{1}{3}}\right),$$

где $S^{*2} = \dfrac{\nu_1 S_1^2 + \nu_1 S_1^2}{\nu_1 + \nu_2}$.

На основе этого можно сформулировать статистику Стьюдента:

$$t_{\nu_1+\nu_2} = \frac{m_1}{S^*\left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right)^{\frac{1}{2}}},$$

необходимую интервальную оценку:

$$\left|t_{\nu_1+\nu_2}\right| \geq t_{1-\frac{\alpha}{2}} \tag{2}$$

и её вероятность наблюдения:

$$P\left\{\left|t_{\nu_1+\nu_2}\right| > t_\alpha\right\} = 2P\left\{\left|t_{\nu_1+\nu_2}\right| > t_{1-\frac{\alpha}{2}}\right\} = \alpha.$$

При больших числах степеней свободы $\nu_1 + \nu_2 > 30$ можно воспользоваться нормальным распределением вместо распределения Стьюдента, т.е. использовать статистику:

$$N = \frac{m_1}{\sqrt{\dfrac{S_{\text{Т}}^2 + S_{\text{Э}}^2}{\nu_1 + \nu_2}}},$$

которая дает более узкий доверительный интервал.

При выполнении условия (2) H_0'' отвергается, в случае альтернативного исхода H_0'' : $m_1 = 0$ принимается.

2. Теперь рассмотрим случай $\sigma^2\left[\Delta P\right] \neq \sigma^2\left[\delta P_\Sigma\right]$.

Рассматриваемая задача сводится к задаче Беренса-Фишера [1] и для её решения не существует безупречного подхода. Приближенное решение этой задачи состоит в том, что для разности матожиданий постулируется значение дисперсии в виде соотношения:

$$S^{*2} = \frac{S_1^2}{\nu_1} + \frac{S_2^2}{\nu_2}.$$

Тогда статистика $y = \frac{m_1}{S^*}$ подчиняется нормальному закону распределения с параметрами N (0,1).

Запишем для сформулированной статистики y интервальную оценку:

$$|y| \geq y_{1-\frac{\alpha}{2}}.$$

При выполнении этого условия гипотеза H_0'' отвергается, в случае альтернативного исхода принимается и, таким образом, принимается сама основная гипотеза.

Для оценки характеристик процесса $\delta P_{\text{М}}(t)$ предлагается метод статистического моделирования. В качестве примера рассмотрим некоторый тепломассообменный аппарат, входными воздействиями для модели которого является внешние тепловые потоки, расходы рабочих тел, а также начальные значения параметров состояния. В качестве выходных параметров можно взять расход, температуру и давление рабочего тела на выходе из аппарата.

Как правило, в ходе физического эксперимента входное воздействие оценивается с большей погрешностью, чем выходные параметры. Часто оценка характеристик этого процесса затруднена и имеются лишь сведения о случайной приборной погрешности. В этом случае, очевидно, можно считать приборную погрешность белым шумом с известной дисперсией.

Статистическая имитация физического эксперимента заключается в многократном математическом моделировании этого эксперимента, при котором начальные условия проведения эксперимента и параметры установки задастся в виде случайных чисел о математическими ожиданиями и дисперсиями, оцененными по результатам их измерений.

Вопросам имитации случайных величин и случайных функций в литературе уделено достаточное внимание.

Для имитации нормального марковского процесса, каковым, например, является временной ряд изменения среднемесячной температуры воздуха, можно воспользоваться зависимостями Т.Андерсона [2], по которым для нормального марковского процесса $\eta(t)$ имеем гауссово распределение с условным средним значением:

$$a\left(\frac{t}{y(S)}\right)=M\left\{\frac{\eta(t)}{y(S)}\right\}=a(t)+k(t,S)\left[k(S,S)\right]^{-1}\left(y(S)-a(S)\right)$$

и условней корреляционной матрицей:

$$k\left(\frac{t}{y(S)}\right)=k(t,t)-k(t,S)\left[k(S,S)\right]^{-1}k(S,t),$$

где $a(t)$ среднее значение процесса $\eta(t)$; $k(t,S)$ – корреляционная функция процесса; $y(S)$ – возможные значений процесса $\eta(t)$.

В результате статистического моделирования, получаем реализации случайного процесса – выходного параметра. Необходимо проверить стационарность и эргодичность расхождения результатов моделирования данных эксперимента.

Стационарность процесса можно проверить построением гистограмм в разных временных точках выходного параметра. Проверка допущения об эргодичности процесса может проводиться по достаточному условию эргодичности стационарной случайной функции, заключающемуся в том, что ее автокорреляционная функция стремится к нулю при $\tau\to\infty$:

$$k_x(\tau)=\frac{1}{n-m}\sum_{i=1}^{n-m}\left(x(t_i)-M_x\right)\left(x(t_{i+m})-M_x\right),$$

где $x(t_i)$ – значение исследуемой функций для *i*-го момента времени;

$$M_x=\sum_{i=1}^{n}\frac{x(t_i)}{n}.$$

Поскольку объем реализаций при статическом моделировании очень велик, обработку материала для получения значений иаточидании $x(t)$ следует организовать рекуррентными методами.

При воспроизведении очередной реализаций «жизни» системы на каж-ом временном шаге также происходит подсчет суммы значений выходного параметра:

$$x_{i\Sigma}(n)=x_{\Sigma}(n-1)+x_i(n);$$

$$x_{i\Sigma^2}(n)=x_{i\Sigma^2}(n-1)+\left(x_i(n)\right)^2.$$

Здесь n – порядковый номер реализации.

После окончания счета всех реализаций при достаточно большом числе реализаций ($n > 40$) можно найти искомые матожидание и дисперсию:

$$\bar{x}_i = \frac{x_\Sigma(n)}{n}; \;\; \sigma_{ix}^2 = \frac{x_{i\Sigma^2}(n)}{n} - \left(x_{i\Sigma}\right)^2.$$

которые затем используются для оценки адекватности модели.

Библиографический список

1. Тутубалин, В.Н. Теория вероятностей / В.Н. Тутубалин. – М.: МГУ, 1972.

2. Полляк, Ю.Г. Вероятностное моделирование на электронных вычислительных машинах / Ю.Г. Полляк. – М.: Сов.радио, 1971.