Тактическое планирование модельного эксперимента. Определение объема статистических испытаний при эксплуатации имитационной модели

Тактическое планирование (ТП) связано с определением эффективности и ресурсоемкости каждого конкретного эксперимента, представляющего собой серию повторяющихся испытаний (прогонов) ИМ, однотипных в смысле задания исходных данных или комбинаций факторов, установленных в ходе СП.

Тактическое планирование сводится к решению двух типов задач:

определение начальных условий в той мере, в какой они влияют на установление стационарного режима работы модели;

снижение погрешности (дисперсии) получаемых при моделировании оценок реакции системы при одновременном сокращении объема испытаний (числа прогнозов).

В отличие от реального объекта в модели всегда существует переходный период, связанный с тем, что случайные процессы, разыгрываемые в модели, требуют определенного времени для выхода в установившийся режим. Возникает задача снижения или исключения влияния начального периода при проведении каждого прогона модели. При этом используют три основных подхода:

увеличение длительности каждого прогона так, чтобы влияние переходного периода было бы заведомо незначительным;

исключение из рассмотрения начального периода (введение этапа предварительной «раскрутки» процесса имитации);

искусственный подбор близких к режимным начальных условий для каждой реализации.

При решении первой задачи тактического планирования в рамках рассмотренных подходов используют, в основном, **эвристические приемы**, опирающиеся на знание физики разыгрываемых в ИМ процессов.

Вторая задача тактического планирования может быть решена строго математически. Она фактически сводится к определению гарантированного объема испытаний (размера выборки, числа прогонов) для получения требуемой точности оценивания компонентов отклика системы, описывающих ее эффективность.

Оценка непрерывнозначной величины. В ходе моделирования в этом случае проводится оценка математического ожидания m и дисперсии D некоторого компонента реакции системы по формулам

$$y_* = \tilde{m} = \frac{1}{n}\sum_{i=1}^{n} u_i , \quad y_{**} = \tilde{D} = \frac{1}{n-1}\sum_{i=1}^{n}(\tilde{m} - u_i)^2 , \qquad (2.5)$$

u_i – наблюдение случайной величины U, полученное в i-ой реализации, $i = \overline{1,n}$;

n – количество повторяющихся испытаний.

Величина $\tilde{m}$, как правило, характеризует такие показатели, как **среднее время выполнения системой своих функций, средний расход энергоресурсов и т.д.**

Величина $\tilde{D}$ – определяет разброс реакции относительно среднего значения, а, кроме того, часто выступает как самостоятельный показатель **точности.**

Для используемых оценок (2.5) выполняются соотношения

$$M[\tilde{m}] = m, \quad \sigma_{\tilde{m}}^2 = M[(\tilde{m} - m)^2] = \frac{D}{n},$$

$$M[\tilde{D}] = D, \quad \sigma_{\tilde{D}}^2 = M[(\tilde{D} - D)^2] = \frac{M_4 - \sigma^4}{n} + \theta(n^{-2}), \tag{2.6}$$

m, $D = \sigma^2$ – истинные математическое ожидание и дисперсия величины U;

$M_4 = M[(U - m)^4]$ – центральный момент U четвертого порядка.

Для гауссовской величины в (2.6) $\sigma_{\tilde{D}}^2 = 2D^2/n$**.**

Метод доверительных интервалов. Необходимо найти такой размер выборки, который гарантировал бы попадание истинных и оцениваемых значений m и/или D внутрь некоторых заранее заданных интервалов с вероятностью $1 - \alpha$.

$$P\left[|\tilde{m} - m| < d_m\right] = 1 - \alpha,$$

$$P\left[\left|\frac{\tilde{D} - D}{D}\right| < d_\sigma\right] = 1 - \alpha, \quad 0 \le d_\sigma \le 1, \tag{2.7}$$

где d_m, d_σ – величины, устанавливающие границы интервалов, которые обычно называют доверительными.

1.Задача оценки среднего. Предположим, сначала, что ошибка оценки отклонения в каждой реализации распределена по гауссовскому или нормальному закону.

Неравенство $|\tilde{m}-m|<d_m$ эквивалентно следующему неравенству:

$$\frac{-d_m\sqrt{n}}{\sqrt{\tilde{D}}}<\frac{(\tilde{m}-m)\sqrt{n}}{\sqrt{\tilde{D}}}<\frac{d_m\sqrt{n}}{\sqrt{\tilde{D}}}. \qquad (2.8)$$

Отсюда следует, что величину доверительного интервала можно определить, используя t-статистику или распределение Стьюдента с k-степенями свободы

$$t=Z\sqrt{k}\big/\sqrt{V},$$

Z – нормально распределенная величина с нулевым средним и единичной дисперсией;

V – независимая от Z случайная величина, имеющая χ^2 распределение с k-степенями свободы;

$V=\sum\limits_{i=1}^{k}V_i^2$, причем V_i – независимые случайные величины, распределенные по нормальному закону $M[V_i\ ]=0$, $M[V_i^2]=1$, $i=\overline{1,k}$.

Величина отношения в (2.8)

$$\Delta m=\frac{(\tilde{m}-m)\sqrt{n}}{\sqrt{\tilde{D}}},$$

подчиняется распределению Стьюдента с $n-1$ степенями свободы.

Доказательство:

$$M[\tilde{m}-m]=0;\ \ M\left[(\tilde{m}-m)^2\right]=\sigma^2\big/n,$$

а величина $\widetilde{D}(n-1)/\sigma^2$ имеет χ^2 распределение с $n-1$ степенями свободы (для разностей $u_i-\widetilde{m}$ в выражении для $\widetilde{D}$ имеется одна линейная связь $\sum\limits_{i=1}^{n}(u_i-\widetilde{m})=0$).

Отсюда получим

$$\frac{(\widetilde{m}-m)}{\sqrt{\dfrac{\sigma^2}{n}}}\times\sqrt{n-1}\times\frac{1}{\sqrt{\dfrac{\widetilde{D}(n-1)}{\sigma^2}}}=\frac{(\widetilde{m}-m)\sqrt{n}}{\sqrt{\widetilde{D}}}.$$

Таким образом, неравенство (2.8) выполняется с вероятностью $1-\alpha$, если

$$P\left[\left|\frac{(\widetilde{m}-m)\sqrt{n}}{\sqrt{\widetilde{D}}}\right|<\frac{d_m\sqrt{n}}{\sqrt{\widetilde{D}}}\right]=S_{n-1}\left(\frac{d_m\sqrt{n}}{\sqrt{\widetilde{D}}}\right)-S_{n-1}\left(-\frac{d\sqrt{n}}{\sqrt{\widetilde{D}}}\right)=$$

$$=2S_{n-1}\left(\frac{d_m\sqrt{n}}{\sqrt{\widetilde{D}}}\right)-1=1-\alpha,$$

где $S_{n-1}(t)$ – функция распределения Стьюдента (t-распределение).

Выражение получено, исходя из следующей цепочки преобразований:

$$S_{n-1}(t)-S_{n-1}(-t)=0{,}5+\int\limits_0^t f(u)du-\left(0{,}5-\int\limits_0^t f(u)du\right)=2\int\limits_0^t f(u)du=$$

$$=2\left(\int\limits_{\infty}^{t} f(u)du-0{,}5\right)=2\int\limits_{-\infty}^{t} f(u)du-1.$$

Отсюда получаем уравнение, связывающее d_m, $\sqrt{\tilde{D}}$ и n

$$\frac{d_m\sqrt{n}}{\sqrt{\tilde{D}}} = t_{kp}(\alpha), \quad n = \frac{t_{kp}^2(\alpha)\tilde{D}}{d_m^2}, \quad t_{kp}(\alpha) = S_{n-1}\left(1-\frac{\alpha}{2}\right),$$

где величина $t_{kp}(\alpha)$ вычисляется по специальным таблицам, исходя из заданного значения α.

Реальное использование полученных соотношений в ИМ основано на проведении **пробной оценки доверительного интервала** или введении правила **автоматического останова** процесса имитации для получения интересующей точности.

При $n > 30$ t-распределение хорошо аппроксимируется гауссовским. При этом

$$S_{n-1}(t) \approx 0{,}5 + \Phi_0(t), \quad \Phi_0(t) = \frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx,$$

$$2\Phi_0(t_{kp}(\alpha)) = 1-\alpha, \quad \frac{d_m\sqrt{n}}{\sqrt{\tilde{D}}} = t_{kp}(\alpha) = \Phi_0^{-1}\left(\frac{1-\alpha}{2}\right).$$

Предположение о нормальности распределения ошибок оценки отклика в каждой реализации может быть неверным.

Тогда используется **неравенство Чебышева**

$$P\left[\left|\tilde{m}-m\right| > d_m\right] = \alpha \le \frac{\sigma_{\tilde{m}}^2}{d_m^2}, \quad \sigma_{\tilde{m}}^2 = \frac{\sigma^2}{n}.$$

Заменяя σ^2 на $\tilde{D}$, получим следующие уравнения:

$$\alpha n d_m^2 = \tilde{D}, \quad n = \tilde{D} / \alpha d_m^2.$$

2.Задача определения дисперсии. Неравенство $\left|\left(\tilde{D}-D\right)/D\right| < d_\sigma$ можно переписать в виде

$$(n-1)(1-d_\sigma) < (n-1)\frac{\tilde{D}}{D} < (n-1)(1+d_\sigma),\ \ 0 < d_\sigma < 1. \qquad (2.9)$$

Величина $(n-1)\dfrac{\tilde{D}}{D}$ подчиняется χ^2 распределению с $n-1$ степенями свободы.

Если n достаточно велико, то ее можно рассматривать как распределенную по нормальному закону с параметрами $m_x = n-1$, $D_x = 2(n-1)$.

Соответственно можно переписать неравенство в виде

$$\frac{-d_\sigma(n-1)}{\sqrt{2(n-1)}} < \frac{\dfrac{(n-1)\tilde{D}}{D}-(n-1)}{\sqrt{2(n-1)}} < \frac{d_\sigma(n-1)}{\sqrt{2(n-1)}},$$

где величина $W = \left[(n-1)\dfrac{\tilde{D}}{D}-(n-1)\right]\Big/\sqrt{2(n-1)}$ распределена по стандартному гауссовскому закону.

Отсюда

$$P\left[\frac{\left|\tilde{D}-D\right|}{D} < d_\sigma\right] = P\left[\left|W\right| < \frac{d_\sigma(n-1)}{\sqrt{2(n-1)}}\right] = 2\Phi_0(t_{kp}(\alpha)) = 1-\alpha.$$

Окончательно получим следующие уравнения:

$$\frac{d_\sigma(n-1)}{\sqrt{2(n-1)}} = t_{kp}(\alpha) = \Phi_0^{-1}\left(\frac{1-\alpha}{2}\right),$$

$$d_\sigma = t_{kp}(\alpha)\sqrt{2(n-1)}, \quad n = 1 + \frac{2t_{kp}^2(\alpha)}{d_\sigma^2}.$$

3.Задача оценка вероятности события. В ходе ИМ оценивают некоторые вероятностные или процентные соотношения для компонентов отклика, описывающих наступление того или иного события (исхода функционирования системы).

Случайная величина U в каждой реализации может принимать два значения: единица с вероятностью «хорошего» исхода p и ноль с вероятностью «плохого» исхода $1-p$. В качестве оценки вероятности p выступает величина

$$y_* = \tilde{p} = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{e}{n},$$

e – количество интересующих исходов в n реализациях.

Требуется найти число испытаний, при котором

$$P\left[\left|\tilde{p}-p\right| < d_p\right] = 1-\alpha, \tag{2.10}$$

d_p – доверительный интервал.

Величина e подчиняется биномиальному распределению

$$e \sim C_n^e p^e (1-p)^{n-e}, \quad C_n^e = \frac{n!}{e!(n-e)!}.$$

При $np > 5$ и $n(1-p) > 5$ или $n\,p(1-p) > 25$ можно пользоваться гауссовской аппроксимацией e с параметрами $m_e = np$ $D_e = np(1-p)$.

Можно переписать исходное неравенство $\left|\tilde{p}-p\right| < d_p$ в виде

$$-\frac{d_p\sqrt{n}}{\sqrt{p(1-p)}} < \frac{e-np}{\sqrt{n}p(1-p)} < \frac{d_p\sqrt{n}}{\sqrt{p(1-p)}}. \tag{2.11}$$

Получим, что соотношение (2.10) эквивалентно следующему:

$$P\left[\left|\tilde{p}-p\right|<d_p\right]=P\left[\left|\frac{e-np}{\sqrt{np(1-p)}}\right|<\frac{d_p\sqrt{n}}{\sqrt{p(1-p)}}\right]=2\Phi_0(t_{kp}(\alpha))=1-\alpha$$

Уравнение, связывающее d_p и n, имеет вид

$$d_p=t_{kp}(\alpha)\sqrt{p(1-p)/n}\,,\quad n=\frac{t_{kp}^2(\alpha)\left(\sqrt{p(1-p)}\right)^2}{d_p^2}.$$

Для того, чтобы избавиться от неизвестного значения p, можно заменить произведение $p(1-p)$ на величину $\max\limits_p\left[p(1-p)\right]=0{,}25$. Тогда получим выражение для гарантированных значений d_p и n

$$d_p^{г}\geq\frac{t_{kp}(\alpha)}{2\sqrt{n}},\quad n^{г}\geq\frac{t_{kp}^2(\alpha)}{4d_p^2}.$$

Таким образом, общий подход к оптимизации процесса имитации в ходе тактического планирования заключается в нахождении уравнений, связывающих доверительные интервалы, то есть требуемую точность оценки величин в ходе ИМ, с необходимым объемом выборки.