Анализ выходных данных для автономной системы

Кирилл Андреев

25 Октября 2011 г.

Наиболее распространенные заблуждения

- Использование единственного прогона модели при анализе систем
- Предположение о независимости случайных величин, получаемых в ходе одного прогона модели

Наиболее распространенные заблуждения

- Использование единственного прогона модели при анализе систем
- Предположение о независимости случайных величин, получаемых в ходе одного прогона модели

Процесс моделирования

Пусть $Y_1, \dots, Y_m$ – стохастический процесс, реализуемый при прогонах модели.

$$\text{Прогоны модели} \left\{ \underbrace{\begin{array}{ccccc} y_{11}, & \dots, & y_{1i}, & \dots, & y_{1m}, \\ y_{21}, & \dots, & y_{2i}, & \dots, & y_{2m}, \\ \vdots & & \vdots & & \vdots \\ y_{n1}, & \dots, & y_{ni}, & \dots, & y_{1m}, \end{array}}_{\text{наблюдения в одном прогоне}} \right.$$

Строки матрицы соответствуют наблюдаемым величинам при различных прогонах модели с использованием различных последовательностей входных случайных чисел.

- Внутри строки величины не независимы
- Внутри столбца величины независимы

Наличие начальных условий

- $Y_1, Y_2, \ldots$ - выходной стохастический процесс
- Функция распределения наблюдаемой $F_i(y|I) = P(Y_i \leq y|I)$зависит от начальных условий I.
- $P(Y_I \leq y)$ – вероятность того, что произойдет событие $\{Y_i \leq y\}$
- $F_i(y|I)$ – переходное распределение выходного процесса в момент времени i при начальных условиях I
  - Функции различны при различных начальных условиях
  - Функции различны в различные моменты времени
- Предполагаем, что существует такое $i \to \infty$, при котором зависимость от начальных условий *практически* исчезает
- Установившееся состояние – <u>распределение</u> наблюдаемой величины перестает зависеть от времени.
- Скорость выхода на установившееся состояние зависит от начальных условий

Выход на стационарное состояние

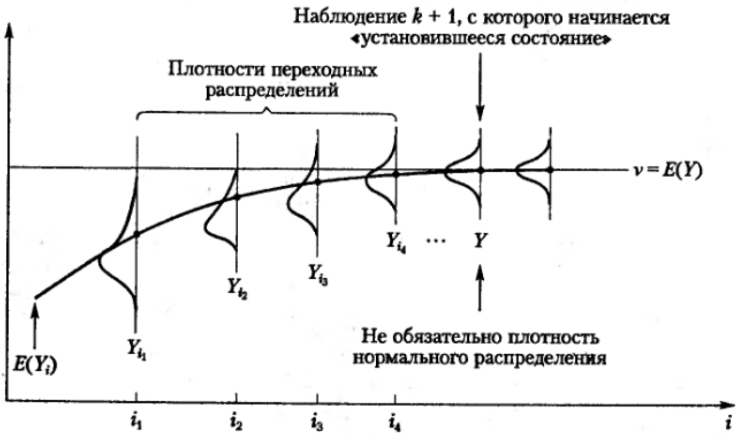


Рис. 9.1. Функции плотности переходного и установившегося распределений для определенного стохастического процесса $Y_1, Y_2, \ldots$ и начальных условий I

Скорость схождения к стационарному состоянию

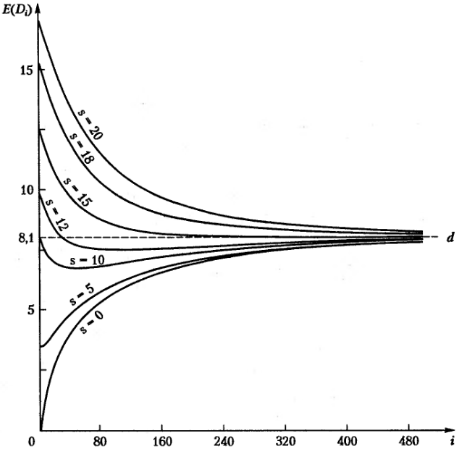


Рис. 9.2. $E(D_i)$ в качестве функции i и число требований в системе s на момент времени 0 для системы массового обслуживания $M/M/1$ с $\rho = 0{,}9$

Типы имитационного моделирования

- Переходный режим имитационного моделирования – наличие терминирующего события. Важен тщательный выбор начальных условий.
- Имитационное моделирование для непереходного режима – невозможность определения терминирующего события, потенциально бесконечное время моделирования.

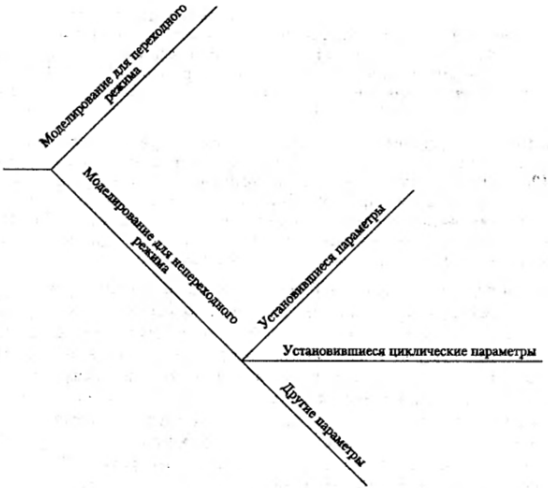
Переходный режим имитационного моделирования

- Моделирование системы с конечным временем работы (рабочий день в банке)
- Моделирование боевых действий – наличие терминирующего события (победа/поражение)
- Моделирование различных производственных циклов

Непереходный режим имитационного моделирования

- Критерий оценки имитационной модели – установившиеся параметры (если существуют) выходного распределения. Характеристики модели при этом считаются неизменными.
- Возможно наличие циклических установившихся параметров: пусть дан стохастический процесс $Y_1, Y_2, \ldots$ непрекращающегося процесса моделирования, в котором нет установившегося состояния. Y_i^C - случайная величина, определенная в $i-$м цикле. Если случайные процессы в каждом цикле сопоставимы, и характеристики этих процессов сходятся, то их называют циклическими установившимися параметрами.
- Если параметры модели изменяются со временем, то возможно свести процесс к переходному.

Типы имитационного моделирования



Оценка средних значений по результатам n прогонов: предположения

- Выполняется n независимых прогонов
- Существует одна оценка искомая оценка критерия (одна величина на один прогон), например среднее ожидание в очереди в банке за день.
- Существует n независимых одинаково распределенных X_j, являющихся искомой оценкой

Оценка средних значений по результатам n прогонов

Процедура с фиксированным объемом выборки: получение приближенного доверительного интервала

Оценка среднего $\mu = E(X)$, где X – оцениваемая случайная величина. Тогда за n прогонов имеем $X_1, \ldots, X_n$ – независимые одинаково распределенные случайные величины. $\overline{X}(n)$ – несмещенная оценка, а $100 - \alpha$-процентный доверительный интервал тогда составит $\overline{X}(n) \pm t_{n-1,1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}}$, где $S^2(n)$ – выборочная дисперсия.

Предположения в процедуре с фиксированным объемом выборки

1. Оцениваемые величины одинаково распределены
2. Оцениваемые величины имеют гауссово распределение

Механизмы получения заданной точности

Сколько необходимо независимых прогонов модели для того, чтобы точно контролировать половину длины доверительного интервала?

- Получение необходимого числа прогонов в предположении "несущественного изменения" выборочной дисперсии $S^2(n)$
- Последовательная процедура получения заданной точности

В дальнейшем зависимость $\overline{X}$ от n исчезает, так как n – случайная величина

Получение числа прогонов при условии несущественного изменения дисперсии

Пусть β – абсолютная погрешность $\overline{X}$, то есть $|\overline{X} - \mu| = \beta$. Тогда доверительный интервал определен как:

$$1 - \alpha \approx P(\overline{X} - \text{половина длины} \leq \mu \leq \overline{X} + \text{половина длины})$$

То есть с вероятностью $\alpha\%$ *среднее* значение наблюдаемой будет лежать внутри доверительного интервала $\pm\beta$. Приближенное необходимое число прогонов составит:

$$n^{\star}_{\alpha}(\beta) = min\left\{ i \geq n : t_{i-1,1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{i}} \leq \beta \right\}$$

По результатам n первых прогонов оцениваем общее необходимое их число и выполняем еще $n - n^{\star}_{\alpha}$ прогонов.

Получение числа прогонов при условии несущественного изменения дисперсии

Пусть необходимо зафиксировать абсолютную погрешность γ, то аналогичным образом можно получить:

$$n_r^{\star}(\gamma) = min\left\{ i \geq n : \frac{t_{i-1,1-\frac{\alpha}{2}}\sqrt{\frac{S^2(n)}{i}}}{|\overline{X}(n)|} \leq \frac{\gamma}{1+\gamma} \right\}$$

Недостатки метода: $\overline{X}(n)$ и $S^2(n)$ не могут быть точными оценками соответствующих параметров генеральной совокупности.

Последовательная процедура получения заданной точности

Необходимо получить оценку μ с заданной относительной точностью $\gamma \in (0, 1)$. Пусть изначально выполнено $n_0 \geq 2$ независимых прогонов модели. Обозначим

$$\delta(n, \alpha) = t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}}$$

1. Выполняем n_0 повторных прогонов модели и задаем $n = n_0$.
2. Вычисляем $\overline{X}(n)$ и $\delta(n, \alpha)$ по $X_1, X_2, \ldots, X_n$
3. Если $\delta(n, \alpha) \leq \overline{X}(n) \cdot \frac{\gamma}{1+\gamma}$, то используем $\overline{X}(n)$ как точечную оценку для μ и останавливаемся или повторяем процедуру иначе, выполняя $n+1$-й прогон модели.

Последовательная процедура получения заданной точности

В результате $\alpha\%$ доверительный интервал с относительной точностью γ составит:

$$I(\alpha, \gamma) = [\overline{X}(n) - \delta(n, \alpha), \overline{X}(n) + \delta(n, \alpha)]$$

Выводы

- Второй способ, как правило, дает существенно большее необходимое число независимых прогонов модели, т.к. на основании малого количества начальных прогонов получаются неточные оценки.
- Обычно следует использовать $n_0 \geq 10, \gamma \leq 0.15$

Общие выводы

- Процедура с фиксированным объемом выборки хороша, когда точность не имеет принципиального значения. Если распределения существенно отличны от нормального, доверительный интервал может оказаться на самом шире, что нежелательно.
- Приблизительная оценка общего количества экспериментов хороша при ограниченных ресурсах на проведение эксперимента.
- Для получения точных оценок лучше пользоваться последовательной процедурой с $n_0 \geq 10, \gamma \leq 0.15$.

Общие выводы

- Никогда не полагаться на результаты, полученные на основании менее пяти прогонов модели.
- Тщательно выбирать оцениваемые параметры. Например - ожидаемая средняя задержка и ожидаемое количество клиентов по времени для одной или нескольких очередей.
- Тщательно подбирать начальные условия. Например – работа банка после обеденного перерыва. Рекомендация: отдельно моделировать процесс выбора начальных данных.

Статистический анализ установившихся параметров: проблема запуска

- Пусть $Y_1, Y_2, \ldots$ – стохастический процесс, полученный в результате прогона модели в непереходном режиме.
- Предположим, что имеется установившееся состояние: $P(Y_i \leq y) = F_i() \to F(y) = P(Y \leq y)$, где Y – установившаяся случайная величина с функцией распределения F.
- ϕ – установившаяся характеристика Y (например, среднее значение)
- Эта оценка отлична от действительной, т.к. мы не можем выбрать "правильные" начальные условия, то есть оценки смещены.

Проблема начального переходного процесса

Пусть необходимо оценить установившееся среднее $\nu = E(Y)$

$$\nu = \lim_{i \to \infty} E(Y_i)$$

Удаление начальных данных, снижающее смещение оценки:

$$\overline{Y}(m, l) = \frac{\sum_{i=l+1}^{m} Y_i}{m - l}$$

Данные, близкие к началу моделирования, могут оказаться *нехарактерными* для рассматриваемой системы.
Проблема начального переходного процесса в выборе l и m таких, чтобы получить необходимое качество получаемой оценки.

Описание метода Велча

1. Выполнение n прогонов модели (более 5!), *достаточно* длительных.
2. Получение на основании нескольких прогонов средних показателей работы системы.
3. Фильтрация высокочастотных колебаний методом оконной фильтрации:

$$\overline{Y}_i(w) = \begin{cases} \frac{\sum_{s=-w}^{w} \overline{Y}_{i+s}}{2w+1} & \text{если } i = w+1, \ldots, m-w \\ \frac{\sum_{s=-(i+1)}^{i-1} \overline{Y}_{i+s}}{2i-1} & \text{если } i = 1, \ldots, w \end{cases}$$

4. Строим график полученного скользящего среднего.

Описание метода Велча

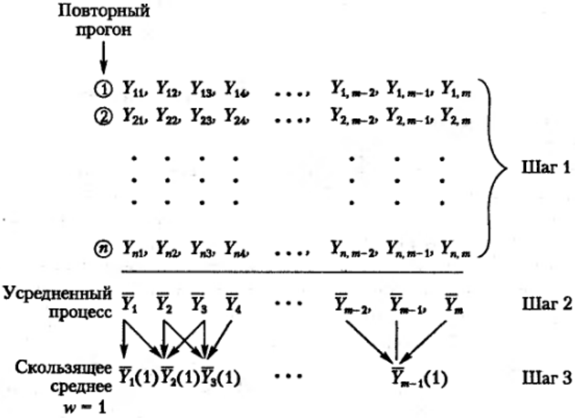


Рис. 9.8. Усредненный процесс и скользящее среднее с $w = 1$, полученные на основании n повторных прогонов имитационной модели продолжительностью m

Общие рекомендации по применению метода

- Выбрать 5 или 10 прогонов, выбрать настолько большое m, насколько это целесообразно в практическом смысле. m должно быть существенно больше ожидаемого значения l.
- Построить график $\overline{Y}_i(w)$ для нескольких значений окна w и выбрать наименьшее значение, при котором соответствующий график будет ”достаточно ровным“ Больший размер окна больше сглаживает получаемый график $\Rightarrow$ слишком большое окно брать не стоит.
- Если не подходит ни одно выбранное значение окна, то стоит выполнить дополнительные прогоны и повторить процедуру.

Общие рекомендации по применению метода

- Существуют более сложные оконные функции для получения результатов (Фильтры Калмана)
- Использование случайной инициализации – позволяет попытаться “угадать” стационарные начальные условия, что увеличивает скорость сходимости.
- Использование данных предыдущих прогонов для инициализации последующих.

Метод репликации и удаления

Пусть Y_{ij} – наблюдаемая на j-м независимом прогоне модели в момент времени i. Пусть X_j определяется следующим образом (среднее по "устойчивым состояниям"):

$$X_j = \frac{\sum_{i=l+1}^{m'} Y_{ji}}{m' - l}$$

X_j – независимые одинаково распределенные случайные величины, $\alpha\%$ доверительный интервал примет вид:

$$\overline{X}(n') \pm t_{n'-1,1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n')}{n}}$$

Метод репликации и удаления

Суть процедуры

1. Выполняется n пробных прогонов модели, по которым определяется длительность переходного периода l.
2. Выполняется еще n' рабочих прогонов, в ходе обработки которых используются $m' - l$ наблюдений.

Рекомендации

- Если в ходе предварительных прогонов модели $l \ll m$, тогда их можно включить в обработку результатов. Неточность метода Велча компенсируется большим объемом данных в "установившемся" состоянии
- Использование двух наборов прогонов модели лучше: не нарушается независимость X_j

Другие методы получения оценки средних

В основном выделяют две стратегии:

- *Процедуры с фиксированным объемом выборки*: выполняется один прогон имитационной модели с произвольной фиксированной длиной, затем по нему производится оценка.
- *Последовательные процедуры*: длина одного прогона имитационной модели последовательно увеличивается.

Методы с фиксированным объемом выборки

- Метод репликации и удаления
- Метод общих средних: Выбор размера групп для получения общих средних
- Авторегрессионный метод: Как выбрать авторегрессионную модель?
- Спектрального анализа: Как выбрать ковариационные интервалы?
- Регенеративный: Что делать, если циклы имеют малую среднюю длину?
- Нормированного временного ряда: Выбор размера групп k.

Множественные оценки показателей работы

Пусть I_s – это $100(1-\alpha)\%$-й интервал для показателей работы μ_s, где $s = 1, 2, \ldots, k$. Тогда

$$P(\mu_s \in I_s \text{ для всех } s = 1, 2, \ldots, k) \geq 1 - \sum_{s=1}^{k} \alpha_k$$

90%-е доверительные интервалы для 10 показателей работы не скажут о нашей системе ничего! Для получения настоящего доверительного интервала нужно использовать 99-процентные интервалы для каждого из показателей работы.