МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ»

А.В. Зюльков, Ю.С. Радченко, А.В. Захаров

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ. ВЕРОЯТНОСТНЫЕ И СТАТИСТИЧЕСКИЕ АСПЕКТЫ

Учебно-методическое пособие для вузов

Воронеж
Издательский дом ВГУ
2017

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

В имитационном моделировании часто возникает необходимость использования методов теории вероятностей и математической статистики. Они необходимы при моделировании входных воздействий на систему и анализе результатов моделирования. При этом нужно особенно внимательно относиться к применению методов статистического анализа, имея в виду зависимость между собой выходных данных при единичном «прогоне» модели.

В пособии кратко

– изложены минимально необходимые сведения из теории вероятностей и математической статистики;
– рассмотрено использование методов статистического анализа на этапе выбора модели входных воздействий;
– приведены задания по моделированию систем с дискретным и непрерывным поведением.

ŋqdbдyeY bƎ dedƎbУзe d eɔɑбɔy ɟqdbУqyè

Числовыми характеристиками распределения случайной величины служат *начальные* m_k и *центральные* μ_k *вероятностные моменты*, задаваемые для дискретных и непрерывных СВ следующими соотношениями

$$E(X^k) \equiv < X^k > \equiv m_k = \sum_{i=1}^{n} x_i^k p_i, \qquad m_k = \int_{-\infty}^{\infty} x^k w(x) dx,$$

$$\mu_k = \sum_{i=1}^{n} (x_i - m_1)^k p_i, \qquad \mu_k = \int_{-\infty}^{\infty} (x - m_1)^k w(x) dx.$$

Здесь $E(\cdot)$ и угловые скобки обозначают усреднение по ансамблю.

Наиболее употребительными характеристиками распределений являются:

1. $E(X) \equiv m_1 = m$ - *математическое ожидание* (*среднее значение*);
2. $Var(X) = E[(X - m_1)^2] \equiv D(X) \equiv \sigma^2 = \mu_2$ - *дисперсия*;
3. $\gamma_1 = \mu_3 / \sigma^3$ - *коэффициент асимметрии* распределения;
4. $\gamma_2 = \mu_4 / \sigma^4 - 3$ - *коэффициент эксцесса* распределения.

В приложениях часто встречаются следующие параметры распределений

– m_0 - *мода* распределения, точка, в которой плотность вероятности $w(x)$ достигает максимума;
– m_e -*медиана* распределения, точка, в которой $F(x) = 0,5$;

Для непрерывных СВ со строго возрастающей функцией распределения $F(x)$ при $0 < q < 1, q$ - *квартиль* $F(x)$ это такое число x_q, для которого $F(x_q) = q$ т.е. $x_q = F^{-1}(q)$. Таким образом, медиана это $x_{0,5}$. Часто используют верхний и нижний квартиль $x_{0,25}$ и $x_{0,75}$. Эти параметры целесообразно использовать в тех случаях, когда моменты распределений не существуют, т.е. интегралы (ряды) расходятся.

Тесноту взаимосвязи СВ X и Y характеризует ковариация $Cov(X,Y) = E[(X - m_X)(Y - m_Y)]$ и корреляция $Cor(X,Y) = Cov(X,Y) / \sigma_X \sigma_Y$.

Чаще всего теоретические распределения задаются в такой форме, что их параметры могут быть отнесены к одному из трех главных типов параметров на основании их физической или геометрической интерпретации, а именно: параметрам положения, масштабным параметрам и параметрам формы.

Параметр положения γ определяет положение области значений распределения на оси абсцисс; обычно γ - средняя (среднее значение m для нормального распределения) или нижняя конечная точка области распределения (в последнем случае параметры положения иногда называют параметрами сдвига). При изменении параметра γ соответствующее ему распределение сдвигается влево или вправо без каких-либо иных изменений. Если распределение случайной величины X имеет параметр положения 0, то распределение случайной величины $Y = X + \gamma$ имеет параметр положения γ.

Масштабный параметр β определяет масштаб (или единицы) измерения значений в диапазоне определения распределения. Стандартное отклонение σ это масштабный параметр для нормального распределения. Изменяя параметр β, соответствующее распределение можно сократить или увеличить без изменения его основной формы. Кроме того, если распределение случайной величины X имеет масштабный параметр 1, то распределение случайной величины $Y=\beta X$ - масштабный параметр β.

Параметр формы α в отличие от параметра положения и масштабного параметра определяет основную форму распределения в общем семействе распределений. Изменение параметра α, как правило, приводит к более фундаментальному изменению свойств распределения, чем изменение параметра положения или масштабного параметра. Некоторые распределения (в частности, экспоненциальное и нормальное) не имеют параметра формы, тогда как другие (например, бета-распределение) могут иметь два таких параметра.

В большинстве имитационных моделей в качестве входных данных используются случайные величины, поэтому выходные данные имитационного моделирования также носят случайный характер. В связи с этим нужно осторожно делать выводы относительно действительных характеристик модели (например, об ожидаемой средней задержке требований в системе массового обслуживанияM|M|1*). Правильное проведение анализа выходных данных невозможно без ознакомления со случайными (стохастическими) процессами.

Стохастический процесс (СП) представляет совокупность «однородных» случайных величин, которые упорядочены во времени и определены в общем, выборочном пространстве. Множество всех возможных значений, которые могут принимать эти случайные величины, называется *пространством состояний*. Если совокупность величин представлена как $X_1, X_2, \ldots$, речь идет о *дискретном* стохастическом процессе, если же как $\{X(t),\ t\ge 0\}$, то о *непрерывном*.

Пример. Рассмотрим систему массового обслуживания с одним устройством обслуживания, например, M|M|1, с независимыми и одинаково распределенными интервалами времени между поступлениями $A_1,\ A_2,\ldots$ независимым и одинаково распределенным временем обслуживания $S_1,\ S_2,\ldots$ и дисциплиной очереди FIFO(first input first output). При генерировании случайных величин $A_1,\ A_2,\ldots$ и $S_1,\ S_2,\ldots$ дискретный стохастический процесс для задержек в очереди ($D_1,\ D_2,\ldots$) может быть определен как

$$D_1=0,\quad D_{i+1}=\max(D_i+S_i-A_{i+1},0),\ \ i=1,2,\ldots$$

Таким образом, моделирование отображает случайные входные величины (например, A_i и S_i) в выходном стохастическом процессе $D_1,\ D_2,\ldots$. Здесь пространство состоя-

* Обозначения (Кендалла) для систем масового обслуживания (СМО) введены далее.

ний представлено множеством неотрицательных вещественных чисел. Обратим внимание, что D_i и D_{i+1} являются положительно **коррелированными**.

Для этой системы $Q(t)$ обозначает количество требований в очереди в момент времени t. Последовательность $\{Q(t),\ t \geq 0\}$ представляет непрерывный стохастический процесс с пространством состояний $\{0, 1, 2, ...\}$.

Для *стационарного в широком смысле* СП $X_1, X_2, ...,$ $m_i = m,$ $\sigma_i^2 = \sigma^2$ для $i = 1, 2...$, а $C_{i,i+j} = Cov(X_i, X_{i+j})$ не зависит от i для $j = 1, 2, ...$. Корреляция между значениями X_i и X_{i+j} не зависит от i

$$\rho_j = \frac{C_{i,i+j}}{\sqrt{\sigma_i^2 \sigma_{i+j}^2}} = \frac{C_j}{\sigma^2} = \frac{C_j}{C_0},\ \ j = 0, 1, 2, ...\ .$$

Пример. Рассмотрим выходной процесс $D_1, D_2, ...$ задержек i -ого требования в очереди из предыдущего примера с коэффициентом загрузки $\rho = \lambda / \omega < 1$(λ это интенсивность поступления требований, а ω — скорость обслуживания). Для этой системы можно вычислить корреляционную функцию ρ_j, график которой при разных значениях коэффициентов загрузки ρ (0,5 и 0,9) изображен на Рис. 5. Из рисунка следует, что значения корреляции ρ_j являются положительными и монотонно снижаются до нуля по мере увеличения j. В частности, $\rho_1 = 0{,}99$ при $\rho = 0{,}9$ и $\rho_1 = 0{,}78$ при $\rho = 0{,}5$. При $\rho = 0{,}9$ $\rho_{50} = 0{,}69$. Таким образом задержки требований в очереди демонстрируют долговременную зависимость. Кроме того, как свидетельствует опыт, выходные процессы для систем массового обслуживания являются положительно коррелированными.

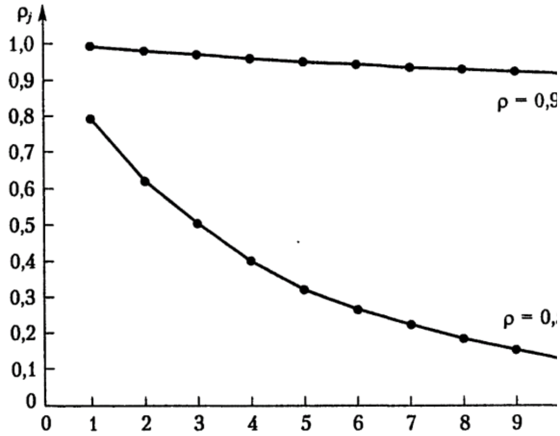

Рис.5

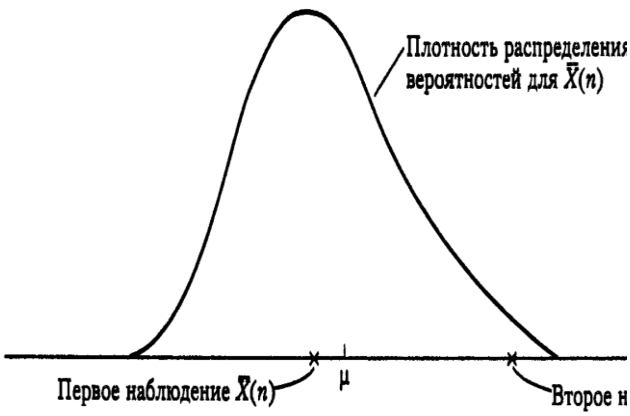

Рис.6

Случайным потоком называется, грубо говоря, последовательность событий во времени, разделенных случайными интервалами. Такой моделью описывается последовательность телефонных вызовов, пакетов данных в системах телекоммуникаций, поток падающих на фотоприемник фотонов и т.д. Математическая теория случайных потоков и связанных с ними точечных случайных процессов хорошо развита.

Простейший случайный поток (модель широко используемая в дальнейшем) представляет собой *ординарную* последовательность независимых событий во времени, разделенных экспоненциально распределенными интервалами. Если параметры λ всех экспоненциальных распределений одинаковы, то имеем стационарный пуассоновский поток с *интенсивностью* λ. Величина λ это среднее число событий потока в единицу времени.

Поток Эрланга r-ого порядка получается путем *просеивания* пуассоновского потока – включения в выходной поток каждой r-ой точки входного (выбрасыванием $r-1$ соседних точек). Поэтому промежуток времени между соседними событиями потока Эрланга представляет собой сумму r независимых экспоненциально-распределенных СВ.

Сведения из математической статистики

Предположим, что $X_1, X_2, ..., X_n$, являются независимыми и одинаково распределенными СВ (наблюдениями) с конечным математическим ожиданием μ и конечной дисперсией генеральной совокупности σ^2. В этом случае *выборочное среднее*

$$\bar{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

является несмещенной (точечной) оценкой $\mu : E[\bar{X}(n)] = \mu$.
Аналогично *выборочная дисперсия*

$$S^2(n) = \frac{1}{n-1}\sum_{i=1}^{n}[X_i - \bar{X}(n)]^2$$

представляет собой несмещенную оценку σ^2, поскольку $E[\mathrm{S}^2(n)] = \sigma^2$. Величины $\bar{X}(n)$ и $S^2(n)$ являются оценками матожидания и дисперсии, поэтому будем обозначать их как как $\hat{\mu}$ и $\hat{\sigma}^2$ соответственно.

Сложность использования $\bar{X}(n)$ в качестве оценки μ без какой-либо дополнительной информации состоит в том, что невозможно определить, насколько близко значение $\bar{X}(n)$ к μ. Поскольку $\bar{X}(n)$ является случайной величиной с дисперсией $Var[\bar{X}(n)]$, в одном эксперименте $\bar{X}(n)$ может быть близко к μ, а в другом их значения могут существенно отличаться (см. Рис.6, для которого принято допущение, что величины X_i являются непрерывными случайными величинами).

Обычно, чтобы оценить точность $\bar{X}(n)$ как оценки μ, необходимо построить доверительный интервал для оценки $\hat{\mu}$. Первым шагом в построении доверительного интервала является оценка дисперсии $Var[\bar{X}(n)]$. Поскольку величины X_i являются независимыми, то справедливо следующее соотношение

$$Var[\bar{X}(n)] = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.$$

Чем больше объем выборки n, тем ближе $\bar{X}(n)$ будет к μ (Рис.6). Кроме того, *несмещенная* оценка дисперсии $Var[\bar{X}(n)]$ может быть получена путем замены σ^2 в предыдущем соотношении на $S^2(n)$

$$Var[\bar{X}(n)] = \frac{1}{n} S^2(n) = \frac{1}{n(n-1)} \sum_{i=1}^{n} [X_i - \bar{X}(n)]^2 \,.$$

Некоторые особенности анализа результатов моделирования. Как свидетельствует опыт, выходные данные моделирования практически всегда являются коррелированными. Следовательно, вышеприведенные сведения о независимых и одинаково распределенных наблюдениях не могут быть непосредственно применены к анализу выходных данных моделирования. Для того чтобы понять опасности обработки выходных данных моделирования, как будто они были независимы, воспользуемся модельюстационарного в широком смысле случайного процесса.

Предположим, что случайные величины $X_1, X_2, \ldots$ получены из такого процесса. Тогда выборочное среднее $\bar{X}(n)$ все еще остается несмещенной оценкой μ, однако выборочная дисперсия $S^2(n)$ более не является несмещенной оценкой σ^2. На самом деле можно доказать, что

$$E[S^2(n)] = \sigma^2 [1 - \frac{2}{n-1} \sum_{j=1}^{n-1} (1 - j/n) \rho_j].$$

Таким образом, если $\rho_j > 0$ (положительная корреляция), как это часто бывает на практике, выборочная дисперсия $S^2(n)$ будет иметь отрицательное смещение: $E[S^2(n)] < \sigma^2$. Этот факт имеет большое значение, поскольку в некоторых программных продуктах имитационного моделирования выборочная дисперсия $S^2(n)$ используется для оценки дисперсии множества выходных данных моделирования, что может привести к существенным ошибкам при анализе.

Рассмотрим оценку дисперсии выборочного среднего $Var[\bar{X}(n)]$, когда величины $X_1, X_2, \ldots$ получены из стационарного в широком смысле СП. Можно утверждать, что

$$Var[\bar{X}(n)] = \frac{\sigma^2}{n} [1 + 2 \sum_{j=1}^{n-1} (1 - j/n) \rho_j].$$

Следовательно, если оценивать $Var[\bar{X}(n)]$ по $S^2(n)$ (что правильно при независимых и одинаково распределенных величинах), как это часто делалось ранее, существует два источника ошибки: смещение в оценке $S^2(n)$, как оценки σ^2 и пренебрежение корреляцией для $E[S^2(n)]$. Фактически, если объединить два предыдущих соотношения получим