К ДИСКУССИИ О КАЧЕСТВЕ ДАТЧИКОВ СЛУЧАЙНЫХ ЧИСЕЛ

В. Н. Задорожный (Омск)

Для реализации случайных факторов, имеющих требуемые вероятностные свойства, в имитационном моделировании (ИМ) используются программные датчики стандартных псевдослучайных чисел (ДСЧ). По назначению любой ДСЧ должен генерировать статистически независимые случайные числа (СЧ), равномерно распределенные на интервале (0, 1). Идеальный датчик (ИД) – это датчик, точно удовлетворяющий требованию равномерности распределения генерируемых на этом интервале СЧ и требованию их независимости. Путем подходящего преобразования идеальных СЧ можно получать идеальные реализации других случайных величин (сл.в.) (или любых случайных объектов), обладающих требуемыми вероятностными характеристиками.

Существует множество статистических тестов, применяемых для проверки равномерности распределения и независимости чисел на выходе ДСЧ. Хорошо известны и методы улучшения свойств недостаточно точных ДСЧ, такие, например, как метод Макларена-Марсальи [1], в котором числа на выходе одного датчика случайным образом перемешиваются с помощью другого. Опубликованная более тридцати лет назад библиография [2] содержит 491 ссылку на литературу по ДСЧ. Однако в книге [3], изданной позднее, Дж. Клейнен приходит к выводу, что «все еще нет высококачественного генератора псевдослучайных чисел». Дискуссионные обсуждения качества ДСЧ не прекращаются и в настоящее время [4, 5]. И чем лучше становятся датчики, тем сложнее выявляются их недостатки, если таковые (достойные внимания) имеются.

Этим определяется актуальность анализа и уточнения показателей, используемых для оценки качества ДСЧ. Ниже рассматриваются некоторые из таких показателей.

Сходимость оценок среднего с ростом числа независимых испытаний

Рассмотрим пример из [5], в котором приводится наблюдаемая при использовании реального датчика зависимость от числа испытаний N погрешности δ расчета числа π методом статистических испытаний через долю точек, попавших во вписанный в квадрат круг (табл. 1). В этом примере погрешность δ при значительном увеличении N часто не снижается, а растет. Так, для N = 20 тыс., N = 50 тыс. и N = 100 тыс. в табл. 1 находим, соответственно, $\delta = 2.2 \cdot 10^{-3}$, $\delta = 2.8 \cdot 10^{-3}$ и $\delta = 8.3 \cdot 10^{-3}$.

Таблица 1

Результаты испытаний ДСЧ из [5]

N	*δ*	*N*	*δ*	*N*	*δ*
1 тыс.	**9.8e-2**	**50 тыс.**	**2.8e-3**	**2 млн.**	**-1.4e-4**
2 тыс.	**8.6e-2**	**100 тыс.**	**8.3e-3**	**5 млн.**	**-1.0e-5**
5 тыс.	**5.6e-3**	**200 тыс.**	**4.3e-3**	**10 млн.**	**2.9e-4**
10 тыс.	**1.8e-2**	**500 тыс.**	**4.3e-3**	**20 млн.**	**4.2e-5**
20 тыс.	**2.2e-3**	**1 млн.**	**2.0e-3**	**50 млн.**	**1.2e-4**

Приведенным примером иллюстрируется разделяемый нами тезис: «реальные датчики равномерных псевдослучайных чисел отнюдь не идеальны» [5]. Попытаемся по возможности точно определить, насколько отличается поведение рассмотренного в этом примере ДСЧ от поведения ИД.

При использовании ИД исходом каждого независимого испытания является попадание разыгранной случайной точки в круг с вероятностью точно $p = \pi/4$, или, с вероятностью $1 - p$, непопадание. По результатам N испытаний вычисляется оценка $\hat{p}$ вероятности p, определяющая оценку $4\hat{p}$ числа π. И поскольку поведение оценки $4\hat{p}$ однозначно определяется поведением оценки $\hat{p}$, вопрос сводится к анализу последней.

Обозначим через z_i индикатор результата i-го испытания: сл.в. z_i принимает значение 1, если в i-м испытании произошло интересующее нас событие, и значение 0 в противном случае. Оценка $\hat{p} = \sum z_i /N$ вероятности p имеет математическое ожидание (м.о.) $\mathrm{M}(\hat{p}) = \mathrm{M}(\sum z_i /N) = p$ и дисперсию $\sigma^2 = \mathrm{D}(\sum z_i /N) = p(1 - p)/N$, а распределение вероятностей сл.в. $\hat{p}$ при больших N с высокой точностью описывается нормальным законом. Ошибка $x = \hat{p} - p$ также распределена нормально с дисперсией $\sigma^2 = p(1 - p)/N$, а ее м.о. $\mathrm{M}(x) = 0$. Это – свойства оценки $\hat{p}$, получаемой при использовании ИД. Исходя из них, можно определить вероятность Q того, что при числе испытаний N_2 получится оценка более точная, чем при числе испытаний N_1. Искомая вероятность $Q = \mathrm{P}(|x_2| < |x_1|)$, где x_2 и x_1 – независимые нормальные сл.в. с нулевыми средними и дисперсиями $\sigma_2^2 = p(1-p)/N_2$ и $\sigma_1^2 = p(1-p)/N_1$ соответственно. Следовательно,

$$Q = P(|x_2|<|x_1|) = P(-x_2 < -x_1, x_2 < 0, x_1 < 0) + \mathrm{P}(-x_2 < x_1, x_2 < 0, x_1 \ge 0) +$$

$$+ \mathrm{P}(x_2 < -x_1, x_2 \ge 0, x_1 < 0) + \mathrm{P}(x_2 < x_1, x_2 \ge 0, x_1 \ge 0) = 4P(x_2 < x_1, x_2 \ge 0, x_1 \ge 0),$$

где все суммируемые вероятности равны в силу соответствующей симметрии плотности двухмерного нормального распределения сл.в. x_2, x_1. Далее, переходя от x_2 и x_1 к центрированным нормированным (стандартным) нормальным сл.в. $\dot{x}_2$ и $\dot{x}_1$, находим:

$$Q = 4\mathrm{P}(\sigma_2\dot{x}_2 < \sigma_1\dot{x}_1,\ \dot{x}_2 \ge 0,\ \dot{x}_1 \ge 0) = 4P(\dot{x}_2 < \frac{\sigma_1}{\sigma_2}\dot{x}_1,\ \dot{x}_2 \ge 0,\ \dot{x}_1 \ge 0) =$$

$$= 4\mathrm{P}(\dot{x}_2 < \sqrt{N_2/N_1}\,\dot{x}_1,\ \dot{x}_2 \ge 0,\ \dot{x}_1 \ge 0) = 4\int_0^\infty \overline{F}_0(k^{-1}t)\, f_0(t)dt\,, \qquad (1)$$

где $k = \sigma_1/\sigma_2 = \sqrt{N_2/N_1}$ – коэффициент сжатия доверительного интервала, f_0 и $\overline{F}_0$ – плотность вероятностей и, соответственно, дополнительная функция распределения стандартной нормальной сл.в. ($\overline{F}_0 = 1 - F_0$).

Коэффициенты сжатия и значения Q, найденные по формуле (1) численным интегрированием, приводятся в табл. 2. Погрешность расчета Q составляет около 0.002.

Таблица 2

Вероятность Q уточнения оценки м.о. в независимых сериях независимых испытаний

N_2/N_1	k	Q	N_2/N_1	k	Q	N_2/N_1	k	Q	N_2/N_1	k	Q
→1	→1	→0.5	5	2.236	0.732	40	6.325	0.900	250	15.81	0.960
2	1.414	0.608	10	3.162	0.805	50	7.071	0.911	400	20	0.968
2.5	1.581	0.641	20	4.472	0.860	100	10	0.937	500	22.36	0.976
4	2	0.705	25	5	0.874	200	14.14	0.955	1000	31.62	0.980

Сопоставим теперь данные табл. 1 и значения Q в табл. 2:

- в табл. 1 имеется 9 случаев удвоения числа опытов, и в 6 из них погрешность снизилась (6 из 9); согласно табл. 2, при удвоении числа испытаний ИД погрешность оценки уменьшается с вероятностью $Q \approx 0.6$;
- из 5 случаев увеличения числа опытов в 2.5 раза в табл. 1 находим 2 или 3 случая снижения погрешности (для ИД при $N_2/N_1 = 2.5$ вероятность $Q \approx 0.64$);
- из 12 случаев увеличения числа опытов вдесятеро в табл. 1 имеется 9 случаев снижения погрешности (для ИД соответствующая вероятность $Q \approx 0.8$);
- погрешность оценки в табл. 1 снизилась во всех 9 случаях увеличения числа опытов в 100 раз (для ИД при $N_2/N_1 = 100$ вероятность этого $Q \approx 0.94$).

Таким образом, характер изменения погрешностей, представленных в табл. 1, вполне отвечает свойствам ИД. Применяя правило «трех сигм», можно видеть, что абсолютная величина погрешностей в табл. 1 также находится в допустимых пределах. Итак, ответ таков: по данным табл. 1 испытанный ДСЧ невозможно отличить от ИД.

На практике, в «штатных» имитационных экспериментах, когда мы увеличиваем число испытаний вдесятеро, то ожидаем снижения погрешности в среднем в $\sqrt{10}$ раз (приблизительно втрое). Соответствующее значение Q в табл. 2 позволяет уточнить наши ожидания: более чем в 20% случаев в подобной ситуации погрешность возрастает (приведенный вывод и расчет Q справедлив для оценок м.о. любых сл.в.).

Формула (1) и табл. 2 получены в предположении, что N_2 новых испытаний независимы от предшествующих N_1 испытаний. Однако при статистическом моделировании мы часто продлеваем единую серию испытаний и наблюдаем за изменением оценки, вычисляемой по всей длине серии. В этом случае N_2 исходов продленной серии испытаний включают в себя N_1 первоначально полученных исходов, и, следовательно, оценки x_2 и x_1 оказываются зависимыми.

Сходимость оценок среднего в продляемой серии независимых испытаний

Пусть по выборке $z_1, \ldots, z_{N1}$ вычисляется оценка $p\square_1$ для м.о. p сл.в. z, и $\mathrm{D}(z) = \sigma^2$. После этого испытания продляются и вычисляется оценка $p\square_2 = (z_1 + \ldots + z_{N1} + \ldots + z_{N2})/N_2$. Найдем вероятность $Q_C = \mathrm{P}(|x_2| < |x_1|)$ того, что погрешность $x_2 = p\square_2 - p$ оценки $p\square_2$ по абсолютной величине будет меньше погрешности $x_1 = p\square_1 - p$ оценки $p\square_1$.

Чтобы перейти от события $|x_2| < |x_1|$, определенного на зависимых нормальных сл.в. x_1 и x_2, к эквивалентному по вероятности событию, определенному на независимых нормальных сл.в., представим $p\square_2$ в виде $p\square_2 = (z_1 + \ldots + z_{N1})/N_2 + (z_{N1+1} + \ldots + z_{N2})/N_2 = \quad = p\square_1 \cdot (N_1/N_2) + p\square' \cdot (N_2 - N_1)/N_2$, где

$\tilde{p}' = (z_{N1+1}+\ldots+ z_{N2})/(N_2 - N_1)$ – независимая «виртуальная» оценка для p. Погрешность x_2 можно тогда выразить через погрешность $x' = \tilde{p}' - p$ «виртуальной» оценки следующим образом:

$$x_2 = \tilde{p}_2 - p =$$

$$= \frac{N_1}{N_2}\tilde{p}_1 + \frac{(N_2 - N_1)}{N_2}\tilde{p}' - p =$$

$$= \frac{N_1}{N_2}(\tilde{p}_1 - p) + \frac{(N_2 - N_1)}{N_2}(\tilde{p}' - p) = \frac{N_1}{N_2}x_1 + \frac{(N_2 - N_1)}{N_2}x', \tag{2}$$

где, как легко видеть, погрешности x_1 и x' являются независимыми нормальными сл.в. с нулевыми м.о. и с дисперсиями σ^2/N_1 и $\sigma^2/(N_2 - N_1)$ соответственно.

С учетом (2) искомая вероятность Q_C может быть представлена в виде:

$$Q_C = \mathrm{P}(|\, x_2\, |<|\, x_1\, |) = P\left(\left|\frac{N_1}{N_2}x_1 + \frac{N_2 - N_1}{N_2}x'\right| <|\, x_1\, |\right). \tag{3}$$

Раскрывая прямые скобки абсолютной величины (путем перебора четырех сочетаний знаков + и –), приводя затем в полученных неравенствах подобные члены и переходя от x_1 и x' к стандартным нормальным сл.в. $\dot{x}_1$ и $\dot{x}'$, сводим (3) к интегралу:

$$Q_C = 2\int_0^\infty \left[F_0\left(\sqrt{k^2-1}\cdot t\right) - F_0\left(-\frac{k^2+1}{\sqrt{k^2-1}}\cdot t\right)\right] f_0(t)dt, \tag{4}$$

где по-прежнему $k = \sqrt{N_2/N_1}$ – коэффициент сжатия доверительного интервала. Значения Q_C, рассчитанные по формуле (4) с погрешностью ≈ 0.002, приведены в табл. 3.

Таблица 3

Вероятность Q_C уточнения оценки м.о. в продляемой серии независимых испытаний

N_2/N_1	k	Q_C	N_2/N_1	k	Q_C	N_2/N_1	k	Q_C	N_2/N_1	k	Q_C
→1	→1	→0.5	5	2.236	0.750	40	6.325	0.901	250	15.81	0.960
2	1.414	0.648	10	3.162	0.813	50	7.071	0.911	400	20	0.968
2.5	1.581	0.675	20	4.472	0.863	100	10	0.937	500	22.36	0.972
4	2	0.727	25	5	0.877	200	14.14	0.955	1000	31.62	0.980

Вероятность уточнения оценки в стратегии продления испытаний (см. табл. 3) немного выше, чем в стратегии их возобновления (см. табл. 2). Отчасти это объясняется ненулевым коэффициентом $r = corr(x_1,x_2)$ корреляции между ошибками x_1 и x_2 в продляемой серии: из соотношения (2) можно найти, что $r^2 = N_1/N_2$. Соответственно остаточная дисперсия $\mathrm{D}(x_2|\, x_1 = t) = \mathrm{D}(x_2)\cdot(1 - r^2)$ оказывается меньше безусловной дисперсии $\mathrm{D}(x_2)$.

Разумеется, маргинальные распределения ошибки x_2 (а следовательно, и безусловные средние значения абсолютной ошибки $|x_2|$) для двух стратегий одинаковы.

Тем не менее стратегия продления серии испытаний экономичнее, так как требует выполнять не N_2, как стратегия возобновления, а только $N_2 - N_1$ новых испытаний.

Используемое выше предположение о нормальном распределении сл.в. x_1, x' и x_2 ограничивает применение полученных результатов условием достаточно больших N_1 и $(N_2 - N_1)$, достигающих хотя бы нескольких десятков или сотен.

Сходимость оценок среднего в серии зависимых испытаний

Зависимые испытания ДСЧ проводятся в [4]: «Для системы массового обслуживания M/M/1 теоретическая средняя длина очереди $q = \rho^2/(1–\rho) = 0.9^2/(1–0.9) = 8.1$. Среднее время ожидания … должно быть равно q/λ и в данном случае ($\lambda = 1$) численно совпадать со средней длиной очереди…». Далее выясняется, что результаты моделирования этой системы на GPSS-W (часть их показана в табл. 4) слишком медленно сходятся к точным значениям. Причиной этого может быть недостаточно высокое качество ДСЧ в используемой версии GPSS-W, проявляемое лишь в достаточно длинных сериях испытаний.

Таблица 4

Результаты [4] испытаний ДСЧ путем моделирования системы M/M/1

Показатель	Теория	Число испытаний		
		50 000	200 000	500 000
Средняя длина очереди	**8.100**	**5.132**	**5.757**	**6.690**
Среднее время ожидания	**8.100**	**5.148**	**5.785**	**6.703**

Определим скорость сходимости оценок, рассматриваемых в табл. 4, с учетом зависимости выборочных значений, по которым эти оценки формируются.

Положительная корреляция между последовательными реализациями $w_1,\ldots,w_N$ времени ожидания w увеличивает дисперсию оценки $\tilde{w} = \sum w_i/N$:

$$D(\tilde{w}) = D\left(\frac{1}{N}\sum_{i=1}^{N} w_i\right) = \frac{1}{N^2}\left[M\left(\sum_{i=1}^{N} w_i\right)^2 - M^2\left(\sum_{i=1}^{N} w_i\right)\right] = \ldots = \frac{\sigma^2}{N}\left(1+2\bar{r}\right), \bar{r} = \frac{1}{N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} r_{ij}\ , \quad (5)$$

где дисперсия σ^2 выборочного значения w_i не зависит от номера заявки i (мы рассматриваем стационарную последовательность реализаций $w_1,\ldots,w_N$), $r_{ij} = corr(w_i, w_j)$.

Значения rij (рассчитанные путем ИМ рассматриваемой системы M/M/1 при длине выборки N = 500 000) убывают по экспоненте с ростом расстояния $s = |i - j|$ между элементами wi и wj выборки. Коэффициент rij существенно зависит от коэффициента загрузки ρ: при ρ = 0.75 он убывает с ростом s на порядок быстрее, чем при ρ = 0.9. Восстановленные для этих ρ по данным ИМ аналитические выражения rij имеют вид:

$$r_{ij} = a_0 e^{-as} = \begin{cases} 0.95e^{-0.0055\cdot|i-j|}, & \rho = 0.9, \\ 0.9e^{-0.0041\cdot|i-j|}, & \rho = 0.75. \end{cases} \quad (6)$$

Подставляя корреляционную функцию $r_{ij} = a_0 e^{-as}$ в (5), и приближая сумму интегралом по $s = |i - j|$, при $N = \infty$ находим $\bar{r} \approx \frac{1}{1 - r(1)} e^{-[1-r(1)]}$. При $r(1) \to 1$, $\bar{r} \approx r(1)/[1–r(1)]$, где $r(1) = r(s)$ при $s = 1$. Вывод последнего приближения для $\bar{r}$ использует грубую аппроксимацию для r_{ij} , определяемую подобранными эмпирически для больших ρ значениями $a_0 \approx 1$ и $a \approx 1 - r(1)$. Используя это приближение для $\bar{r}$ в (5), далее получаем:

$$D(\hat{w}) = \frac{\sigma^2}{N}\left(1 + 2\bar{r}\right) \approx \frac{\sigma^2}{N}\left(\frac{1 + r(1)}{1 - r(1)}\right),\ \rho \to 1,\ N \to \infty. \qquad (7)$$

При $\rho = 0.9$ найденные путем ИМ параметры σ^2 (дисперсия времени ожидания) и $r(1)$ равны приблизительно 82 и 0.991 соответственно (точное значение $\sigma^2 = 80.19$). Из (7) для N из табл. 4 находим: при N = 50 тыс. $D(\hat{w}) \approx 0.6^2$, при N = 200 тыс. $D(\hat{w}) \approx 0.3^2$, и при N = 500 тыс. $D(\hat{w}) \approx 0.2^2$. Отклонения на «три сигмы» оценки $\hat{w}$ от среднего $M(w) = 8.1$ при этих N составляют соответственно $3{\cdot}0.6 = 1.8$, $3{\cdot}0.3 = 0.9$ и $3{\cdot}0.2 = 0.6$. Если же $D(\hat{w})$ рассчитывать непосредственно по формуле (5), не используя грубую аппроксимацию (7), то получим $\bar{r} \approx 70$ и для N = 50 тыс., 200 тыс. и 500 тыс. – допустимые отклонения 1.4, 0.7 и 0.5 соответственно. В табл. 4 отклонения оценок $\hat{w}$ от 8.100 примерно вдвое превышают эти допуски. Таким образом, расчеты подтверждают качественную оценку [4]: «результаты, связанные с ожиданием, оставляют желать лучшего».

Применение рассмотренного теста к студенческой версии 4.3.2 GPSS World (copyright 2000) при тех же N, что перечислены в табл. 4, дает приемлемые результаты – например, 8.702, 8.362 и 8.727 – соответственно этим N (правда, результат 8.727 уже лежит «на грани допустимого»). Тест был продлен; результаты более длинных прогонов модели с фактическими и допустимыми погрешностями представлены в табл. 5.

Таблица 5

Результаты моделирования системы M/M/1 на GPSS World версии 4.3.2

Показатель	Число испытаний				
	1 млн.	10 млн.	100 млн.	1 млрд.	2 млрд.
Оценка среднего времени ожидания	8.332	8.061	8.087	8.043	7.799
Фактическая погрешность оценки	0.232	0.039	0.013	**0.057**	**0.301**
Допустимая погрешность («три сигмы»)	0.404	0.064	0.040	**0.006**	**0.004**

При N = 1 млрд. погрешность результатов моделирования на порядок превысила допустимую, а при N = 2 млрд. – почти на два порядка. Приведенные в табл. 5 допустимые погрешности требуют дальнейшего уточнения, так как они рассчитаны по $r(1)$ (и с учетом $\bar{r}$), оцененного с помощью проверяемого ДСЧ; правда, для этого использовалась хорошо испытанная часть генерируемых им чисел ($N \leq 500\,000$). В любом случае, проверяемый датчик либо хорош только на указанных недлинных выборках, либо не пригоден и на них. С другой стороны, рассматриваемый метод тестирования также требует дальнейшего уточнения, например, путем аналитического или численного расчета коэффициента $1 + 2\bar{r}$ в соответствии с (5) или путем его более точного расчета методом ИМ с использованием нескольких хорошо проверенных датчиков, реализованных разными языками программирования.

Выводы

Найденные вероятности Q и Q_C уточнения оценок среднего развивают идею тестирования реальных ДСЧ, предложенную и проиллюстрированную в [5], и позволяют перевести ее на язык проверки статистических гипотез. Эти вероятности полезно учитывать и для корректной интерпретации результатов «штатных» статистических экспериментов. Формулы (5)–(7) для расчета дисперсии D(w□) оценки w□ среднего времени ожидания (с учетом корреляции элементов выборки) уточняют методику тестирования ДСЧ, используемую в [4], позволяют корректно строить доверительные интервалы, накрывающие точное значение M(w), и могут быть встроены в пакеты ИМ.

Литература

1. **Харин Ю. С., Степанова М.** Д. Практикум на ЭВМ по математической статистике: Для мат. спец. ун-тов. Минск, 1987. 304 с.
2. **Nance R. E. and Overstreet C.** (1972). A bibliography on random number generation // Computing Rev., 13. P.495-508.
3. **Клейнен Дж.** Статистические методы в имитационном моделировании : Пер. с англ./ Под ред. Ю.П.Адлера и В.Н. Варыгина. М.: Статистика, 1978. Вып. 1. 221 с. Вып. 2. 335 с.
4. Оценка системы моделирования GPSS World. Режим доступа: http://gpss.ru/paper/ryzhikov/index_w.html.
5. **Рыжиков Ю. И., Соколов Б. В., Юсупов Р. М.** Проблемы теории и практики имитационного моделирования // Третья всероссийская конференция «Имитационное моделирование. Теория и практика. ИММОД-2007». Т. 1. СПб.: ФГУП ЦНИИТС, 2007. С. 58–70.