Лекция 14.

Моделирование случайных величин

Датчики случайных чисел

Пусть $\eta \in U(0,1)$

Def. Члены последовательности $y_1, y_2, \dots, y_n$, которые можно рассматривать как экспериментальные данные случайной величины η, называются псевдослучайными числами. А устройство или алгоритмы для их получения - датчики (или генераторы) случайных чисел

I. Физические датчики

Примерами физических датчиков может быть секундомер, случайной величиной может быть миллисекунды случайно остановленного секундомер, или европейская рулетка с 36 ячейками для шарика

Простейшим способом случайное число можно сгенерировать, подбрасывая монету, где за одну сторону принимается 1, а за другую - 0

Th. Случайная величина $\eta \in U(0,1) \iff$ разряды ξ_i в ее двоичной записи $\sum\limits_{i=1}^{n} 2^{-1}\xi_i$ имеют распределение Бернулли $B_{\frac{1}{2}}$

Если требуется точность 2^{-n}, то бросать монету нужно n раз

Физические датчики довольно примитивными, однако их значения сложно передавать компьютеру. Также две последовательности случайных чисел могут отличаться при одинаковых условиях

Знаменитый пример использования физических датчиков - использование лавовых ламп компанией Cloudflare для генерации чисел для использования в шифровании

II. Таблицы случайных чисел

Пусть имеется таблица псевдослучайных чисел - результат работы некоторого датчика. Случайным образом выбиралась строка и столбец и, начиная с этого места, выбиралась последовательность случайных чисел

Такой способ использовался до широкого появления компьютеров и сейчас устарел

III. Математические датчики

Обычно математический датчик - это рекуррентная последовательность вида $y_n = f(y_{n-1})$, однако способ генерации может быть значительно сложнее

В качестве математического датчика разберем мультипликативный датчик: задается большое число m, начальное число k_0 и множитель a, при этом m, k_0 и a - взаимно простые

Последовательность y_n задается по формуле:
$\begin{cases} k_n \equiv ak_{n-1} \pmod m & \text{- остаток от деления } ak_{n-1} \text{ на } m \\ y_n = \frac{k_n}{m} \in (0,1) \end{cases}$
Есть рекомендация использовать $m = 2^{31} - 1 = 2147483647$, $a = 630360016$ или 764261123 (алгоритм Дж. Фишмана и Л. Мура)

Позднее предложили такой датчик (алгоритм Вичмена-Хилла):
$a_1 = 171$, $m_1 = 30269$, $a_2 = 172$, $m_2 = 30307$, $a_3 = 170$, $m_3 = 30323$
Члены y_n', y_n'', y_n''' задаются как мультипликативные датчики, а итоговое значение вычисляется как дробная часть от их суммы: $y_n = \{y_n' + y_n'' + y_n'''\}$
Преимущества: работает быстрее (числа меньше), период датчика - $3 \cdot 10^{13}$, а алгоритма Фишмана и Мура - $2 \cdot 10^9$
Наблюдателю кажется, что, чем сложнее алгоритм датчика, тем более случайным он кажется

Моделирование непрерывного распределения

Квантильное преобразование (или метод обратной функции)

На курсе теории вероятности выражали такую теорему:

Th. Пусть $F(x)$ - непрерывная, строго возрастающая функция распределения
Если случайная величина $\eta \in U(0,1)$, то $\xi = F^{-1}(\eta)$ имеет функцию распределения $F(x)$

Ex. Показательное распределение E_α: $F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\alpha x}, & x \geq 0 \end{cases}$
Обратной к ней функцией будет $x = -\dfrac{1}{\alpha}\ln(1-y) \Longrightarrow \xi = -\dfrac{1}{\alpha}\ln\eta \in E_\alpha$

Ex. Нормальное распределение
Если $\xi \in N(0,1)$, то $\sigma\xi + a \in N(a, \sigma^2)$, поэтому достаточно уметь моделировать стандартное нормальное распределение
$F(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{z^2}{2}}\, dz$
Если $\eta \in U(0,1)$, то $\xi = F^{-1}(\eta) \in N(0,1)$ (`НОРМ.СТ.ОБР` в Excel)

Nota. Этот алгоритм простой и универсальный, но не самый эффективный

Нормальные случайные числа

I. На основе ЦПТ

Пусть $\eta_i \in U(0,1)$. Тогда $E\eta = \frac{1}{2}$, $D\eta = \frac{1}{12}$, $S_n = \eta_1 + \dots + \eta_n$ и по ЦПТ:

$\frac{S_n - nE\eta}{\sqrt{nD\eta}} = \frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \rightrightarrows N(0,1)$

Уже при $n = 12$ получается неплохое приближение $S_{12} - 6 \approx N(0,1)$

II. Точное моделирование пары независимых значений $N(0,1)$

> Пусть $\eta_1, \eta_2 \in U(0,1)$ и независимы. Тогда случайные величины $X, Y \in N(0,1)$ и независимы:
> $X = \sqrt{-2\ln\eta_1}\cos(2\pi\eta_2)$
> $Y = \sqrt{-2\ln\eta_1}\sin(2\pi\eta_2)$

> Пусть $X, Y \in N(0,1)$ независимы, тогда плотность $f_{X,Y}(x,y) = f_X(x) f_Y(y) = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}$
> Перейдем к полярным координатам:
> $x = r\cos\varphi, y = r\sin\varphi, |J| = r$
> Плотность: $f_{\Phi,R}(\varphi, r) = \frac{1}{2\pi}e^{-\frac{1}{2}r^2}r$
> Так как $\varphi \in U(0, 2\pi)$, то $f_\Phi(\varphi) = \frac{1}{2\pi}$ и $f_{\Phi,R}(\varphi,r) = \underbrace{\frac{1}{2\pi}}_{f_\Phi} \cdot \underbrace{re^{-\frac{1}{2}r^2}}_{f_R} \Longrightarrow \Phi$ и R - независимы
> Смоделируем f_R методом обратной функции
> $F_R(r) = \int_0^r re^{-\frac{r^2}{2}}dr = -\int_0^r re^{-\frac{r^2}{2}}d\left(-\frac{r^2}{2}\right) = e^{-\frac{r^2}{2}}\Big|_0^r = 1 - e^{-\frac{r^2}{2}} = \eta_1$. Тогда $r = \sqrt{-2\ln\eta_1}$
> Аналогично $F_\Phi(\varphi) = \int_0^\varphi \frac{1}{2\pi}d\varphi = \frac{\varphi}{2\pi} = \eta_2 \Longrightarrow \varphi = 2\pi\eta_2$

Такой датчик использует всего три сложных операции (логарифм, корень и косинус) и моделирует нормальное распределение очень точно

Быстрый показательный датчик

> **Th.** Пусть независимая случайная величина $\eta_1, \eta_2, \dots, \eta_n, \eta_{n+1}, \dots, \eta_{2n-1} \in U(0,1)$, $\xi_1, \xi_2, \dots, \xi_{n-1}$ - упорядоченные значения случайных величин $\eta_{n+1}, \dots, \eta_{2n-1}$, $\xi_0 = 0$, $\xi_n = 1$
> Тогда случайная величина $\mu_i = -\frac{1}{\alpha}(\xi_i - \xi_{i-1})\ln(\eta_1 \cdot \eta_2 \cdot \dots \cdot \eta_n) \in E_\alpha$ и независимы, $1 \le i \le n$

Экономим на вычислении значения (считаем логарифм один раз), но теряем при сортировке.
Алгоритм оптимален при $n = 3$: $\eta_1, \eta_2, \eta_3, \eta_4, \eta_5 \in U(0,1)$ и $\eta_4 < \eta_5$, тогда $\mu_1 = -\frac{1}{\alpha}\eta_4\ln(\eta_1\eta_2\eta_3)$; $\mu_2 =$

$-\frac{1}{\alpha}(\eta_5-\eta_4)\ln(\eta_1\eta_2\eta_3)$; $\mu_3=-\frac{1}{\alpha}(1-\eta_5)\ln(\eta_1\eta_2\eta_3)$

Моделирование дискретных случайных величин

I. Общий метод (или квантильное преобразование)

Пусть ξ - дискретная случайная величина с законом распределения $P(\xi=x_i)=p_i$

Разбиваем единичный отрезок на отрезки длин p_1, p_2 и так далее

Пусть $r_m=\sum\limits_{i=1}^{m} P_i$ - границы отрезков

Если $\eta_i\in U(0,1)$ и $\eta_i\in[r_{i-1},r_i)$, то $\xi_i=x_i$

В частности, так можно смоделировать распределение Бернулли: делим отрезок на две части; если $\eta_i<1-p$, то $\xi_i=0$, если $\eta_i\geq 1-p$, то $\xi_i=1$

II. Биномиальное распределение

$B_{n,p}: P(\xi=k)=C_n^k p^k q^{n-k}$, $k=0,1,\ldots,n$ - число успехов при n экспериментах

Берем n значений датчика $y_1,\ldots,y_n\in U(0,1)$. Если $y_i<1-p$, то $z_i=0$, если $y_i\geq 1-p$, то $z_i=1$

Тогда $\xi_i=\sum\limits_{i=1}^{n} z_i\in B(n,p)$

III. Геометрическое распределение

$G_p: P(\xi=k)=pq^{k-1}$ - номер первого успеха в испытании

Берем серию значений датчика $y_i\in U(0,1)$ до тех пор, пока не будет $y_i\geq 1-p$. Тогда $\xi=i$ - номер эксперимента

IV. Распределение Пуассона

$\Pi_\lambda: P(\xi=k)=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,\ldots$

Используем тот факт, что распределения показательное и Пуассона довольно тесно связаны

> **Th.** Пусть $\mu_1,\mu_2,\ldots,\mu_n$ - независимые случайные величины с распределением E_λ
> Пусть $S_n=\mu_1+\cdots+\mu_n, N=\max n$ такое, что $S_N\in[0,1]$
> Тогда $N\in\Pi_\lambda$

Алгоритм: проводим серию k-ую серию испытаний до тех пор, пока $\prod\limits_{i=1}^{k} y_i<e^{-\lambda}$, тогда $\xi_i=k$