

第三节 随机变量的抽样

根据随机变量遵循的分布规律使用一种方法获取它的具体值,叫做抽样(sampling).随机变量的抽样方法很多,不同的分布采用的方法不尽相同.区间[0,1]上均匀分布的随机变量的抽样是最基本的方法.在计算机上,其他各种分布的随机抽样方法都是它的基础上产生的.

计算机上产生的随机数是按照确定的算法产生的,它遵循一定的规律,显然不是真正随机的,因此这种随机数我们叫做伪随机数(random number).只要伪随机数能通过一系列的统计检验,就可以把它们当作真正的随机数放心地使用,而不会引起太大的误差.

产生均匀分布的伪随机数的常用方法有平方取中法、线性同余法和广义同余法等.由于目前计算机上常用的高级语言(如C,Pascal,Fortran等)都有产生均匀分布随机数的系统函数,我们可以直接使用而不必关心其实现原理,在此不作专门介绍.首先介绍MATLAB软件中产生区间[0, 1]上的均匀分布随机数的系统函数R=rand(n),它产生 $n \times n$ 阶均匀随机矩阵.一般R=rand(m,n),产生 $m \times n$ 阶均匀随机矩阵.

下面简单介绍由均匀随机数以及某种算法产生其他分布随机数的方法.用 r_1, r_2, \dots 表示独立同分布于U[0, 1]的随机数列.

直接抽样法

设连续的随机变量X有分布函数

$$F(x) = \int_{-\infty}^x f(t)dt \text{ 其中 } f(t) \text{ 是 } X \text{ 的密度函数。}$$

因为 $0 < F(x) < 1$, 令 $r = F(x)$, 若函数 $F(x)$ 的反函数存在, 那么对随机变量X的抽样可由公式 $x = F^{-1}(r)$ 产生出.

例子

近似抽样法

近似抽样方法的步骤是设一组独立同分布的随机变量 $r_1, r_2, \dots, r_n, \dots$ 利用中心极限定理

$$\frac{\frac{1}{n} \sum_{i=1}^n r_i - Er_i}{\sqrt{Dr_i/n}} \sim N(0,1), \quad \text{当 } n \rightarrow \infty \text{ 时}$$

设 r_1, r_2, \dots 为[0,1]区间上的均匀随机数列, 则 $Er_i = 1/2, Dr_i = 1/12$, 从而有, 当n充分大时,

$$\frac{\frac{1}{n} \sum_{i=1}^n r_i - \frac{1}{2}}{\sqrt{\frac{1}{12n}}} \sim N(0,1)$$

常用的是 $n=12$ 的情形, $u_{12} = \sum_{i=1}^{12} r_i - 6 = \sum_{i=1}^6 r_{2i} - r_{2i-1}$, 其中 u_{12} 是标准正态分布的随机数, 即由12个区间[0,1]上均匀随机数产生一个标准正态随机数. 如果还要产生一般正态分布的随机数, 则只需变换 $z = \sigma u + \alpha$. 另外, 还有一种利用二维正态变换产生的标准正态分布的随机数, 其变换如下:

$$\begin{cases} x = \sqrt{-2 \ln r_1} \cos 2\pi r_2 \\ y = \sqrt{-2 \ln r_1} \sin 2\pi r_2 \end{cases} \quad \text{当 } r_1 \neq 0 \text{ 时}$$

MATLAB 中各种分布下产生随机数的命令

常见的分布函数

均匀分布 U[0,1]

MATLAB 语句

R=rand(m,n)

均匀分布 $U[a,b]$

指数分布 $E(\lambda)$

正态分布 $N(\mu, \sigma)$

二项分布 $B(n, p)$

了松分布 $P(\lambda)$

$R = \text{unifrnd}(a, b, m, n)$

$R = \text{exprnd}(\lambda, m, n)$

$R = \text{normrnd}(\mu, \sigma, m, n)$

$R = \text{binornd}(n, p, m, n1)$

$R = \text{poissrnd}(\lambda, m, n)$

以上语句均产生 $m \times n$ 的矩阵.