BITS Pilani
Pilani Campus

MATH F113
Probability and Statistics

Dr. Shivi Agarwal
Department of Mathematics

Simulation

- To generate a data of values of a random variable, we need to perform the random experiment. (several times for large data).
- Such data can be used to infer more about the random variable, like parameters.
- Sometimes even though *we know the density*, information about r.v. which is not mathematically obtainable can be inferred from data.

Simulation is probably the most important technique used in analysing a number of complex systems in estimating their characteristics. This technique is used when the analytic methods of analysing a complex system are not known. Whenever mathematical methods for getting the characteristics of a complex system do not exist, simulation method enables one to obtain approximate values of these characteristics

Basically, simulation is a technique of manipulating a model of a system through a process of imitation. In simulation, the characteristics of a system are estimated by artificially doing a large number of sampling observations on the variables of the system following the conditions of the system. Thus, simulation is essentially the technique of conducting a large number of sampling experiments on the model of the system

Simulation allows us to generate the values of the random variables without actually performing the random experiment.

The data can be used for estimating a parameter or other information about r.v. which is only approximate.

The generation of a sample from a r.v. X with known density

Monte Carlo Simulation

The principle behind the Monte Carlo simulation technique is representative of the given system under analysis by a system by some known probability distribution and their drawing random samples from probability distribution by means of random numbers. In case it is not possible to describe a system in terms of standard probability distribution such as normal, binomial, Poisson, Exponential etc. an empirical probability distribution can be constructed.

The Monte Carlo simulation technique consists of following steps:

(i) Setting up a probability distribution for variables to be analysed.

(ii) Building a cumulative probability distribution for each random variable.

(iii) Random numbers: Assign an appropriate set of random numbers to represent value of range (interval) of values for each random variable.

iv) Conduct the simulation experiment by means of random sampling.

(v) Repeat step (iv) until the required number of simulation runs has been generated.

(v) Design the implement a course of action and maintain control.

The process depends on first generating random digits.

This can be done using computer programs, or tables, or some other suitable method.
Example : To generate a digit randomly, put 10 balls with distinct labels 0, 1, ---, 9 in a bucket and pick one ball out of them *randomly*, read label on it.

Each digit must have same probability of being chosen.

One can also generate 2-digit random numbers by repeating the same experiment twice.

Random Digit

- The digits from 0,1,2,3,.....,9 are arranged in such a manner that while choosing, all digits are equally likely, thus

$$f(x) = \frac{1}{10} \quad for \quad x = 0,1,2,3,....,9$$

One can also use 3 digit random numbers or 4 digit random numbers; where say 3 digit random numbers are 000, 001, ..., 999(0 must be written even if it is the starting digit and no decimal signs; they are not real numbers), each with probability $\frac{1}{1000}$

Rather than performing this experiment, we may use random number tables readily available. (p.693-694)

Random Number Table

13962	70992	65172	28053	02190	83634	66012	70305	66761	88344
43905	46941	72300	11641	43548	30455	07686	31840	03261	89139
00504	48658	38051	59408	16508	82979	92002	63606	41078	86326
61274	57238	47267	35303	29066	02140	60867	39847	50968	96719
43753	21159	16239	50595	62509	61207	86816	29902	23395	72640
83503	51662	21636	68192	84294	38754	84755	34053	94582	29215
36807	71420	35804	44862	23577	79551	42003	58684	09271	68396
19110	55680	18792	41487	16614	83053	00812	16749	45347	88199
82615	86984	93290	87971	60022	35415	20852	02909	99476	45568
05621	26584	36493	63013	68181	57702	49510	75304	38724	15712

To choose a 1 digit random number, look at the 1st digit occurring in the table.
Alternately we can also randomly pick a location (row and column number).
Pick the number in table at that location.

To produce a sequence of 10 *one-digit* random numbers, pick randomly a location. Start reading a sequence from there continuously row-wise or column-wise. If you come to an end of a row or column, continue with the 1st entry of the next row or column.

Two produce a sequence of 10 r-digit random numbers from a randomly chosen position, start reading r consecutive numbers in the table. Again can continue row-wise or column-wise as in 1-digit case.

- To generate a sample from a distribution, look at probability of its outcomes. For each outcome, assign a group of random numbers which has same probability as that outcome such that different groups are disjoint.

- Read random numbers from table (or given otherwise), see if they belong to any group. Take the corresponding outcome.

Assigning of group can be done in several ways

- Example: To simulate 5 flips of a fair coins, can divide random digits 0,1,2,...,9 in 2 groups, each of probability ½, eg. 0,1,...,4 and 5, 6, .., 9 or even digits & odd digits. Assign one group to head and other to tail, say

  0,2,4,6,8-----tail

  1,3,5,7,9-----head

Use of cumulative probabilities : If the outcomes of random expt are real numbers, **cumulative probabilities** are *commonly used* to form groups of random numbers associated to outcomes.

- If the probabilities are given upto 2 (respectively 3) decimal places we can divide 2 digit (respectively 3 digit) random numbers in groups corresponding to outcomes using cumulative probabilities.

Simulation of known density.

We will discuss simulation for

- Discrete r.v
- Continuous r.v.

Discrete density

P[X=a]=u is a number between 0 and 1.

•a $\longrightarrow$ A group of random numbers whose probability is u (equivalent group)

•The groups allocated with different values of a should be disjoint.

•To use r-digit random numbers, the probabilities u taken upto r decimal places

•the groups of r-digit random numbers are disjoint and exhaustive.

Method of allocation of groups:

- Generally c.d.f.s are used for allocation of a group random numbers to values of X.

Let $x_1 < x_2 < \ldots, x_n < \ldots$ be all values of X.

To use r-digit random numbers,

- calculate upto r decimal places $F(x_1), \ldots, F(x_n), \ldots$
- $x_i \longrightarrow$ those r-digit random numbers such that after placing a decimal point before them we get a real number in the interval $[F(x_{i-1}), F(x_i))$.

(For allocation to x_1, we may start at r-digit random number 00…0)

Exercise: Probability that a computer software salesperson will make 0,1,2,3,4 or 5 sales on a day are 0.10, 0.30, 0.25, 0.15, 0.14 and 0.06. Use 2 digit random numbers 15, 45, 23, 72, 90 to simulate his sales on 5 days. Hence estimate his expected daily sale.

- Allocation of random numbers :

Value of Random var.	Prob	Cumulative prob	Allocation of two digit Random no.s
0	0.10	0.10	00-09
1	0.30	0.40	10-39
2	0.25	0.65	40-64
3	0.15	0.80	65-79
4	0.14	0.94	80-93
5	0.06	1	94-99

- In the exercise, given random numbers correspond to outcomes 1,2,1,3,4.

Estimated expected sale= (1+2+1+3+4)/5=2.2

2-digit random no.	Value of random var.
15	1
45	2
23	1
72	3
90	4

Generating Discrete Random Variables:

Ex: Let X be a discrete random variable with the following distribution.

X	1	2	3	4
P(X)	0.2	0.1	0.4	0.3

Step1: Compute the CDF:

$$F(x)=\begin{cases} 0, & x<1 \\ 0.2, & 1\le x<2 \\ 0.3, & 2\le x<3 \\ 0.7, & 3\le x<4 \\ 1, & x\ge 4 \end{cases}$$

Now, given a random observation from the uniform distribution, called a random number, a value of X can be generated as follows:

$$
\begin{aligned}
\text{Set } X &= 1, \text{ if } u<0.2 \\
&= 2, \text{ if } 0.2\le u<0.3 \\
&= 3, \text{ if } 0.3\le u<0.7 \\
&= 4, \text{ if } u\ge 0.7
\end{aligned}
$$

It can be seen that:

$P[X=1] = P[u<0.2] = 0.2$

$P[X=2] = P[0.2 \le u < 0.3] = 0.1$

$P[X=3] = P[0.3 \le u < 0.7] = 0.4$

$P[X=4] = P[0.7 \le u < 1] = 0.3$

Note: In practice we use random numbers having finite number of digits. Suppose we pick up two digit random numbers .00,.01,.02……99.

If a number u is picked up from these 100, Then

$P[u<0.2] = 0.2 = 20/100$, since the two digit random numbers are .00, .01…0.19 (i.e. 20 in number), then we write it as:

$X=1$ if $u<0.20$

Generation of Bernoulli random variate

Let X is a Bernoulli random variable with parameter p $(0 < p < 1)$. The point probability function of X is given by

$$p(x) = P(X = x) = p^x (1-p)^{1-x}, \qquad x = 0,1$$

The distribution function of X is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ 1-p, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

If u is a random number and uniformly distributed over [0,1), then set

$$x = 0, \quad \text{if } u < 1 - p$$

$$= 1, \quad \text{if } u \geq 1 - p$$

where x is the desired value of X.

Generation of Binomial random variate

Let X is a Binomial random variable with parameters n and p (0 < p < 1). The point probability function of X is given by

$$p(x) = P(X = x) = \binom{n}{x} p^{x} (1-p)^{1-x}, \quad x = 0, 1, 2, \ldots, n$$

By using cdf, we can simulate *x*.

Generation of Binomial random variate

Let X be a binomial random variable with parameter n and p. Then X can be expressed as the sum of n independent Bernoulli random variables with parameter p each, i.e.

$$X=X_1+X_2+\ldots\ldots\ldots\ldots+X_n$$

where X_1, X_2, …X_n are independent Bernoulli random variables each having parameter p.

Hence a value x of X can be generated by generating a value x_i of $X_{i,}$ $i = 1, 2, \ldots, n$, and then writing

$$x = x_1 + x_2 + \ldots + x_{n.}$$

Algorithm for generating a binomial variate

Step 1: Set $x = 0$

Step 2: For $i = 1$ to n do:

Pick a random number u_i

If $u_i < 1 - p$, set $b = 0$

otherwise set $b = 1$

$x = x + b$

x is the required generated value of X.

Ex: Using three digit random numbers 775, 130, 382, 086, 895 generate one observation of the binomial random variable X with $n = 5$ and $p = 0.4$.

Sol: Step 1: $x = 0$

Step 2: $i = 1$, $u_1 = 0.775 > 1 - p = 0.6$, $b = 1$

$$x = x + b = 0 + 1 = 1$$

$i = 2$, $u_2 = 0.130 < 1 - p = 0.6$, $b = 0$

$x = x + b = 1 + 0 = 1$

Do it for all given random numbers….

Finally $x = 2$.

Ex: Using three digit random number 775, 130, 382, 683 generate four observations of the binomial random variable X with n = 5 and p = 0.4 through cdf of X.

Sol:

X	0	1	2	3	4	5
F(X)	0.078	0.337	0.683	0.913	0.990	1.00

Simulation of continuous random variable.

This is based on the following

Fact : If X is a continuous r.v. and F(x) is the c.d.f. of X then U=F(X) is a random variable having uniform distribution on [0,1).

Def: Given a value u in [0,1) of the uniform r.v. U on [0,1), a value x of X can be generated as follows:

$$u = F(x) \quad for\ all\ x \Rightarrow$$

$$x = F^{-1}(u)$$

This is called inverse transformation method.

Example: A continuous r.v X has the pdf:

$$f(x) = \begin{cases} x-1, \ 1 \le x \le 2 \\ 3-x, \ 2 \le x \le 3 \\ 0, \ \ elsewhere \end{cases}$$

Generate a random sample of size 4 by using two digit random numbers 24, 94, 50 and 36

Generation of Exponential random variate

Let X have an exponential distribution with parameter β. The distribution function of X is given by

$$F(x)=1 - e^{-x/\beta}, \quad x > 0$$

Let u be a value of U ~UD [0,1), we set $1 - e^{-x/\beta} = u$

i.e.

$$x = -\beta \ln(1-u)$$

Thus if β=2 then the 2-digit random numbers 54 and 09 generate values

$$-2\ln(1-0.54) \text{ and } -2\ln(1-0.09) \text{ resp.}$$

Note: U is uniformly distributed on (0,1) iff 1 - U is uniformly distributed on (0,1). Therefore,

$$x = -\beta \ln(u), \quad u \neq 0$$

can be used for generating a value x of X.

Generation of Uniform random variate

Let X be a random variable which is uniform distributed on (a,b). The distribution function of X is given by

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a < x < b \\ 1, & x > b \end{cases}$$

Let u be a value of uniform r.v. U on [0,1), then the corresponding value x of the uniform random variable on (a, b) is given by

$$\frac{x-a}{b-a} = u$$

or, $$x = a + (b-a)u$$

To generate values of uniform random variable X on (4, 7) corresponding to 3-digit random numbers 235, 789, 178; note that cdf of X is

$$F(x) = \begin{cases} 0; & \text{for } x < 4 \\ \dfrac{x-4}{3} & \text{for } 4 \le x \le 7 \\ 1 & \text{for } x > 7. \end{cases}$$

Thus we can tabulate values of X generated thus :

3-digit random no.	u	$X=F^{-1}(u)$
235	0.235	3(0.235)+4
789	0.789	3(0.789)+4
178	0.178	3(0.178)+4

Example: A continuous r.v X has the pdf:

$$f(x) = \begin{cases} -x^3, -1 < x < 0 \\ \dfrac{1}{2}x,\ 0 < x < 1 \\ \dfrac{1}{2},\ 1 < x < 2 \\ 0,\ \text{elsewhere} \end{cases}$$

Compute its cdf:

$$F(x) = \begin{cases} 0, x < -1 \\ \frac{1}{4}(1 - x^4), -1 \le x < 0 \\ \frac{1}{4}(1 + x^2), 0 \le x < 1 \\ \frac{1}{2}x, 1 \le x < 2 \end{cases}$$

$$\text{Also}, F(0) = 1/4 = 0.25, F(1) = 1/2 = 0.5$$
$$F(2) = 1$$

Let F(x) = u where u be a value in [0,1) of U, uniform r.v. on [0,1)

$$If\ 0 \le u < 0.25, Then\ x = -(1-4u)^{\frac{1}{4}}$$

$$If\ \ 0.25 \le u < 0.5, Then\ x = (4u-1)^{\frac{1}{2}}$$

$$If 0.5 \le u < 1,\ \ then \quad x = 2u$$

The values u of U~UD[0,1) can be generated from the r-digit random numbers just by placing a decimal point before it. Then value of X can be found. For example two digit random number say 37, corresponds to 0.37, then

$X = (4(0.37)-1)^{1/2}$ etc.