Acceptance-Rejection Technique

- The efficiency of the technique depends on being able to minimize the number of rejections.
- This technique is normally used when cumulative density function cannot be easily obtained.

Acceptance-Rejection Technique

Example: use following steps to generate uniformly distributed random numbers between 1/4 and 1

Step 1.

Generate a random number R

Step 2a.

If $R \geq 1/4$, accept X = R, goto Step 3

Step 2b.

If $R < 1/4$, reject R, return to Step 1

Step 3.

If another uniform random variate on [1/4, 1] is needed, repeat the procedure beginning at Step 1. Otherwise stop.

Acceptance-Rejection Technique

Do we know if the random variate generated using above methods is indeed uniformly distributed over [1/4, 1]? The answer is Yes. To prove this, use the definition. Take any $1/4 \leq a < b \leq 1$

$$P(a < R \leq b | 1/4 \leq R \leq 1) = \frac{P(a < R \leq b)}{P(1/4 \leq R \leq 1)} = \frac{b-a}{3/4}$$

which is the correct probability for a uniform distribution on [1/4,1].

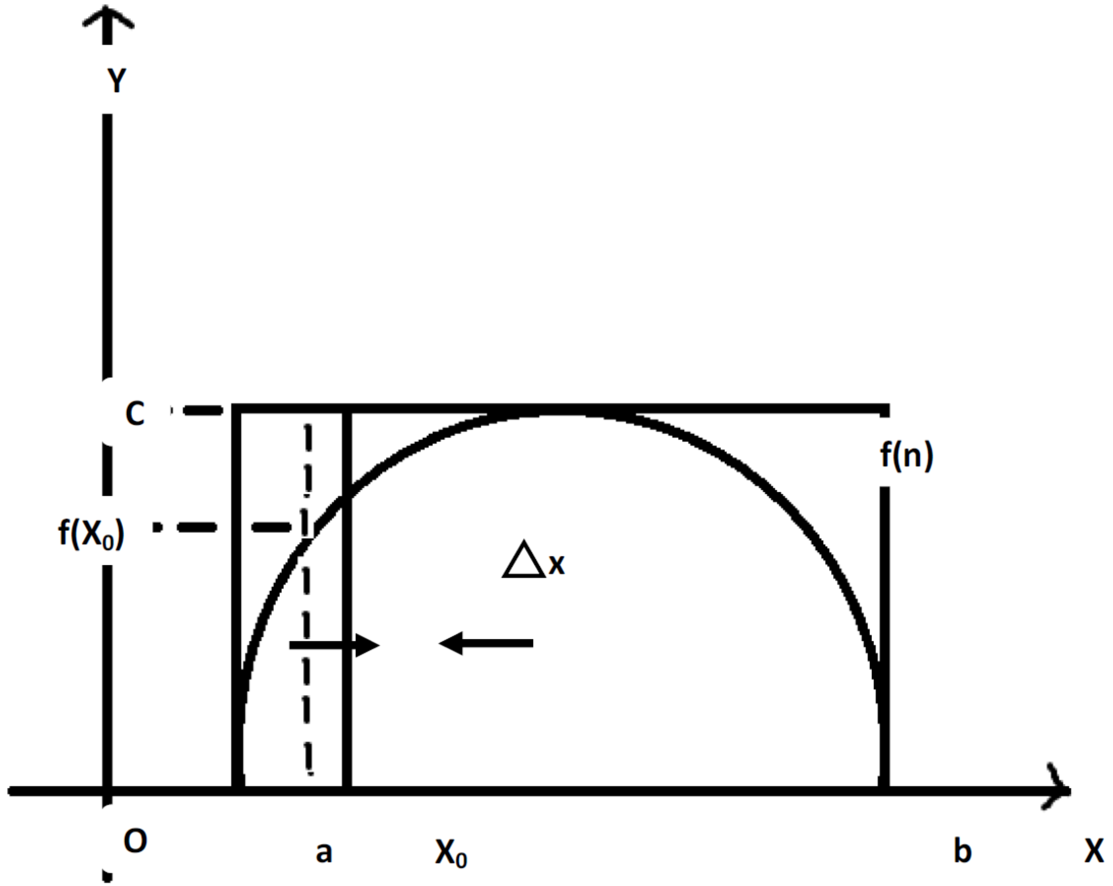
REJECTION METHOD

This method is applicable when the probability density function f(x) has a lower and upper limit, its range ‘a’ and ‘b’ respectively and an upper bound ‘c’.

Steps:

- Compute the values of two independent uniformly distributed variate v1 and v2.
- Compute X0= a+v1(b-a)
- Compute Y0=cv2
- If Y0 ≤ f(X0), accept X0 as the desired output otherwise repeat the process with two new variants.

REJECTION METHOD



REJECTION METHOD

- This method is closely related to Monte-Carlo technique. Here the probability density function is enclosed in a rectangle with sides '(b-a)' and 'c'.
- The first 3 steps of the method just create a random point and the last step relates the point to the curve of the pdf.
- If the points fall on or below the curve, the value 'X0' is accepted as a sample from the desired distribution, otherwise the point is rejected and the process is repeated.
- As the curve is a pdf, so the area under the curve must be '1' it means that the scale of the graph must be chosen so that the area of the rectangle must be 1. i.e. (b-a)c= 1

REJECTION METHOD

- Consider a small interval of ‘X’ axis ‘Δx’ at the point X0.
- The probability of ‘X’ being less than or equal to ‘X0’ is given by the cdf

$$P(X \leq X0) = cdf = F(X0)$$

- The probability of ‘Y0’ falling on or below the curve to the left of ‘X0’ is the ratio of the area under that part of the curve to the rectangle with sides (X0-a) and ‘c’.
- Since ‘X0’ is uniformly distributed between ‘a’ and ‘b’ the probability that ‘X’ will be in the range

$$\boxed{P(a \leq X \leq X_0) = X_0 - a/b - a}$$

DISADVANTAGES

- Two uniform variants must be calculated for each trail point and since some points are rejected more than two uniform variants are needed for creation of each output point.
- The correct application of the rejection method requires that the pdf been limited it means the function should be identically ‘0’ below a and above b.

Test for Random Numbers

The algorithms of testing a random number generator are based on some statistics theory, i.e. testing the hypotheses.

The basic ideas are the following, using testing of uniformity as an example.

We have two hypotheses,

H_0 : Random number generator is uniformly distributed (Known as *null hypothesis)*

H_1 : Random number generator is not uniformly distributed (Known as alternative *hypothesis)*

We are interested in testing result of $H_{0,}$, reject it, or fail to reject it .

Test for Random Numbers (cont.)

- To see why we don't say *accept H null*
- *Accepting H null* mean that the distribution is truly uniform.
- But this is impossible to state, without exhaustive test of a *real* random generator with infinite number of cases.
- So we can only say *failure to reject H null*, which means no evidence of non-uniformity has been detected on the basis of the test.
- This can be described by the saying ``so far so good''.
- On the other hand, if we have found evidence that the random number generator is not uniform, we can simply say *reject H null*.

Test for Random Numbers (cont.)

- It is always possible that the H_0 is true, but we rejected it because a sample landed in the H_1 region, leading us to reject H_0 This is known as ***Type I* error.**
- Similarly if H_0 is false, but we didn't reject it, this also results in an error, known as ***Type II* error.**

	Null hypothesis is TRUE	**Null hypothesis is FALSE**
Reject null hypothesis	Type I Error (False positive)	Correct outcome! (True positive)
Fail to reject null hypothesis	Correct outcome! (True negative)	Type II Error (False negative)

Test for Random Numbers (cont.)

- With these information, how do we state the result of a test?
- A level of statistical significance **α** has to be given.
- The level **α** is the probability of rejecting the H null while the H null is true (thus, Type I error).

$$\boldsymbol{\alpha = P\,(\text{reject } H_0 \mid H_0 \text{ True})}$$

- The probability Should be as little as possible.
- Typical values are 0.01 (1%) or 0.05 (5%).
- Decreasing the probability of Type I error will increase the probability of Type II error.
- We should try to strike a balance.

Test for Random Numbers (cont.)

- Two categories:
  - Testing for uniformity:

$$H_0: \quad R_i \sim U[0,1]$$

$$H_1: \quad R_i \nsim U[0,1]$$

    - Failure to reject the null hypothesis, H_0, means that evidence of non-uniformity has not been detected.
  - Testing for independence:

$$H_0: \quad R_i \sim \text{independently}$$

$$H_1: \quad R_i \nsim \text{independently}$$

    - Failure to reject the null hypothesis, H_0, means that evidence of dependence has not been detected.

Test for Random Numbers (cont.)

- Number of tests are performed to check the uniformity and independence of random numbers
- Two types of tests are
- **<u>Frequency test</u>**: compares the distribution of the set of numbers generated to a uniform distribution. Few are:
  - Kolmogorov-Smirnov Test
  - Chi-square Test

Both tests measure the agreement between the distribution of a sample of generated random numbers and the theoretical uniform distribution.

Test for Random Numbers (cont.)

Both tests are based on the null hypothesis of no significant difference between the sample distribution and the theoretical distribution.

- **Autocorrelation test :** tests the correlation between the two numbers and compares the sample correlation to the desired correlation, zero
  - Runs test
  - Gap test
  - Pokers test

Kolmogorov-Smirnov Test(K-S Test)

- Used as a test of goodness of fit
- Ideal when the size of the sample is small.
- It compares the cumulative distribution function for a variable with a specified distribution.
- The null hypothesis assumes no difference between the observed and theoretical distribution.

Kolmogorov-Smirnov Test(K-S Test)

- The value of test statistic 'D' is calculated as:

$$D = \text{Maximum}|Fo(X) - Fr(X)|$$

Fo(X)= Observed cumulative frequency distribution of a random sample of n observations.

Fo(X)=k/n = (No. of observations ≤ X)/(Total no. of observations).

Fr(X) = The theoretical frequency distribution.

The critical value of D is found from the K-S table values for one sample test.

Kolmogorov-Smirnov Test(K-S Test)

- **Acceptance Criteria:** If calculated value is less than critical value accept null hypothesis.
- **Rejection Criteria:** If calculated value is greater than table value reject null hypothesis.

Kolmogorov-Smirnov Test(K-S Test)

- In a study done from various streams of a college 60 students, with equal number of students drawn from each stream, are interviewed and their intention to join the Drama Club of college was noted.

	B.Sc.	B.A.	B.Com	M.A.	M.Com
No. in each class	5	9	11	16	19

It was expected that 12 students from each class would join the Drama Club. Using the K-S test to find if there is any difference among student classes with regard to their intention of joining the Drama Club.

Kolmogorov-Smirnov Test(K-S Test)

- Ho: There is no difference among students of different streams with respect to their intention of joining the drama club.
- We develop the cumulative frequencies for observed and theoretical distributions.

Kolmogorov-Smirnov Test(K-S Test)

Streams	No. of students interested in joining		$F_O(X)$	$F_T(X)$	$\lvert F_O(X)-F_T(X)\rvert$
	Observed (O)	Theoretical (T)			
B.Sc.	5	12	5/60	12/60	7/60
B.A.	9	12	14/60	24/60	10/60
B.COM.	11	12	25/60	36/60	11/60
M.A.	16	12	41/60	48/60	7/60
M.COM.	19	12	60/60	60/60	0
Total	n=60				

Kolmogorov-Smirnov Test(K-S Test)

Test statistic $|D|$ is calculated as:

$D = \text{Maximum}|F_0(X) - F_T(X)|$

$= 11/60 = 0.183$

The table value of D at 5% significance level is given by

$D_{0.05} = 1.36/\sqrt{n}$

$= 1.36/\sqrt{60}$

$= 0.175$

Since the calculated value is greater than the critical value, hence we reject the null hypothesis and conclude that there is a difference among students of different streams in their intention of joining the Club.

Kolmogorov-Smirnov Test

- The test consists of the following steps
  - **Step 1:** Rank the data from smallest to largest
    $R_{(1)} \leq R_{(2)} \leq \ldots \leq R_{(N)}$
  - **Step 2:** Compute

$$D^{+} = \max_{1 \leq i \leq N} \left\{ \frac{i}{N} - R_{(i)} \right\}$$

$$D^{-} = \max_{1 \leq i \leq N} \left\{ R_{(i)} - \frac{i-1}{N} \right\}$$

  - **Step 3:** Compute $D = \max(D^{+}, D^{-})$
  - **Step 4:** Get D_{α} for the significance level α
  - **Step 5:** If $D \leq D_{\alpha}$ accept, otherwise reject H_0

Kolmogorov-Smirnov Critical Values

Degrees of Freedom (N)	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.828
4	0.564	0.624	0.733
5	0.510	$\boxed{0.565}$	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.410	0.490
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	0.325	0.361	0.433
14	0.314	0.349	0.418
15	0.304	0.338	0.404
16	0.295	0.328	0.392
17	0.286	0.318	0.381
18	0.278	0.309	0.371
19	0.272	0.301	0.363
20	0.264	0.294	0.356
25	0.24	0.27	0.32
30	0.22	0.24	0.29
35	0.21	0.23	0.27
Over 35	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

Kolmogorov-Smirnov Test

- Example: Suppose N=5 numbers: 0.44, 0.81, 0.14, 0.05, 0.93.

Step 1:

Step 2:

i	1	2	3	4	5
$R_{(i)}$	0.05	0.14	0.44	0.81	0.93
i/N	0.20	0.40	0.60	0.80	1.00
$i/N - R_{(i)}$	0.15	0.26	0.16	-	0.07
$R_{(i)} - (i\text{-}1)/N$	0.05	-	0.04	0.21	0.13

Arrange $R_{(i)}$ from smallest to largest

$D^+ = max\{i/N - R_{(i)}\}$

$D^- = max\{R_{(i)} - (i\text{-}1)/N\}$

Step 3: $D = \max(D^+, D^-) = 0.26$

Step 4: For $\alpha = 0.05$,

$D_\alpha = 0.565 > D = 0.26$

Hence, H_0 is not rejected.