Lecture # 9

AMCS 241 - STAT 250

Chapter 4: Simulation of RVs

4.1 Introduction

. In many fields of science and engineering, computer simulations are used to study random phenomena and the performance of an engineering system in a noisy or uncertain environment due to analytic intractability of some complex proba models.

. The heart of these applications is that it is possible to simulate a RV

4.2 Generation of RV Numbers

. There are several algorithms to generate samples from $U[0,1]$ and we will assume in this lecture that these algorithms are available to us in the form of routines/commands in our standard Math Software packages (MATLAB, MATHEMATICA, ...)

4.3 Method of Inverse Transform

θ_ Method

- Suppose we want to generate a RV X with a prescribed CDF $F_X(x)$.

- ~~Notice~~ Note that the RV

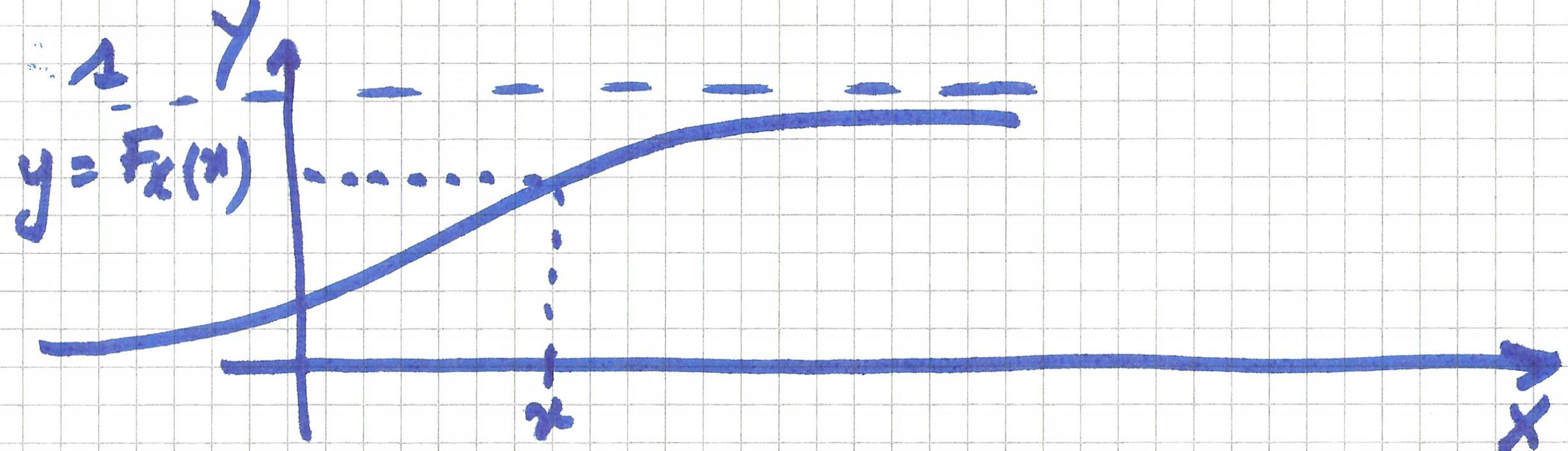
$$Y = F_X(X) \text{ is } U[0,1]$$

because

$$f_Y(y) = \frac{f_X(x)}{|F_X'(x)|} = 1$$

$$\text{''} \quad f_X(x)$$

- Thus given $U[0,1]$ random number generator Y, the inverse transform $X = F_X^{-1}(Y)$ will have the desired CDF $F_X(x)$.

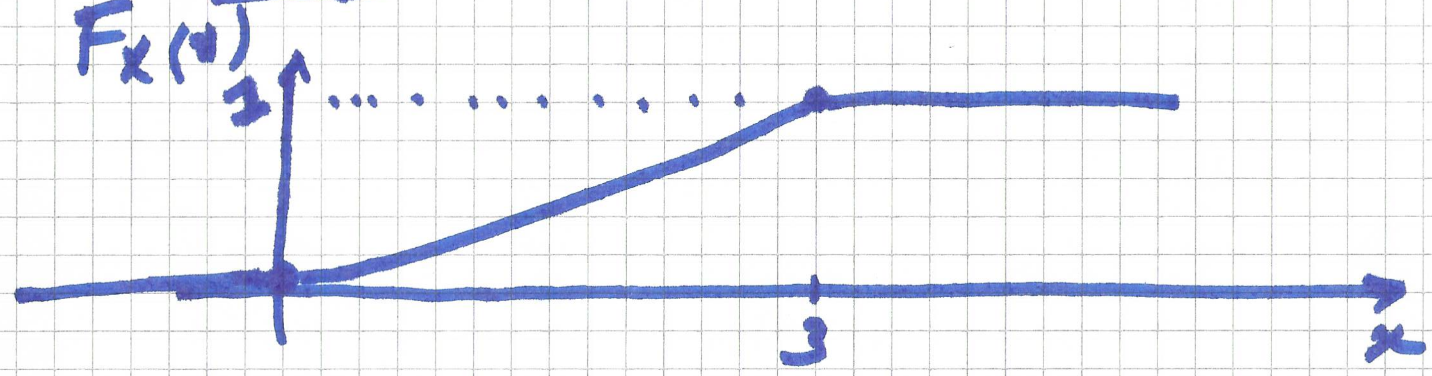


The algorithmic steps for the inverse transform method are as follows

1. Generate a random number from $Y \sim U[0,1]$. Call it y
2. Compute the value x such that $F_X(x) = y$ (i.e $x = F_X^{-1}(y)$)
3. Take x to be the random number generated.

b - Examples

1



$$F_X(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \dfrac{x^2}{9} & \text{for } 0 \leq x \leq 3 \\ 1 & \text{for } x \geq 3 \end{cases}$$

$$F_X(x) = y \iff \frac{1}{9} x^2 = y$$

$$\cancel{y = \mp\sqrt{9x}} \qquad x = \mp\sqrt{9y}$$

$$Y \sim U[0,1] \longrightarrow X = +\sqrt{9y} \sim F_X(x).$$

(2) Example 2

Suppose we want to generate a RV with an exponential distribution given by

$$f_X(x) = \lambda e^{-\lambda x} u(x) \quad \text{with } \lambda > 0$$

$u(x)$ ↑ unit step function

$$F_X(x) = \int_0^x \lambda e^{-\lambda t} dt$$

$$= \frac{\lambda}{-\lambda} e^{-\lambda t} \Big|_0^x = -(e^{-\lambda x} - 1)$$

So $$F_X(x) = 1 - e^{-\lambda x}$$

Therefore given y, we can get x by the mapping

$$1 - e^{-\lambda x} = y$$

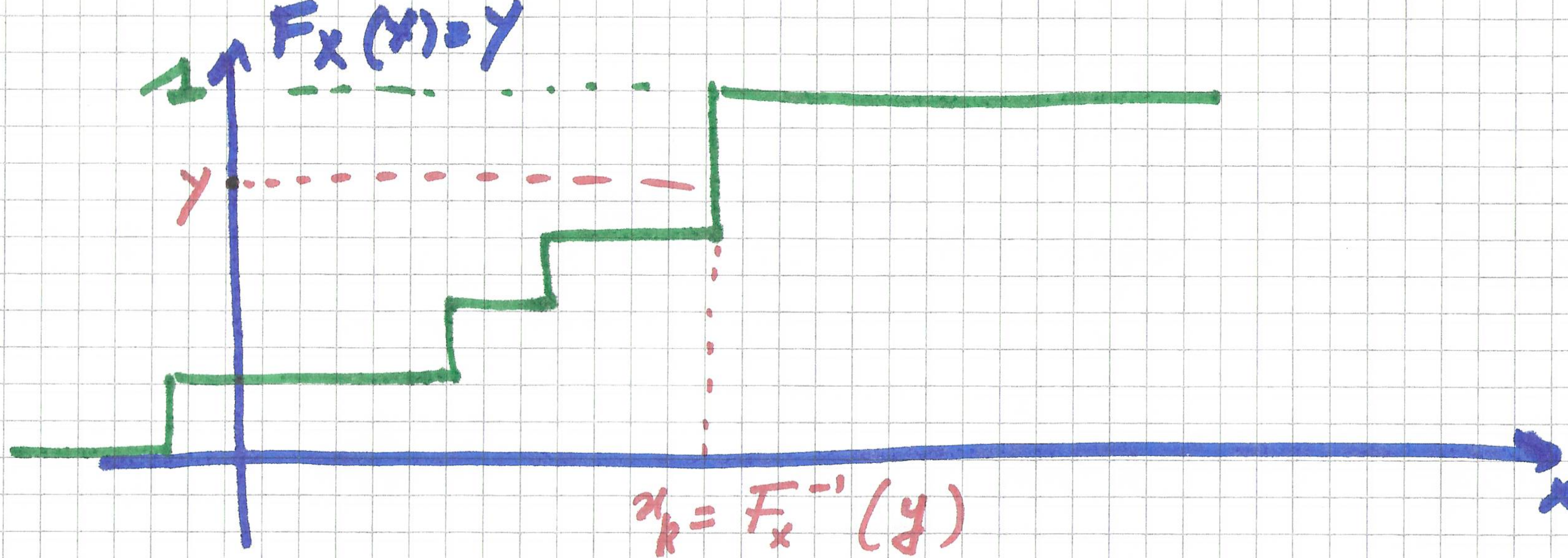
$$\vdots$$

$$x = -\frac{\ln(1 - \overset{U[0,1]}{y})}{\lambda}$$

Since $1-y$ is also uniformly distributed over $[0,1]$. Thus the above expression can also be re-written as

$$\boxed{x = -\frac{\ln y}{\lambda}}$$

Example 3 : Generation of discrete RV



recall $P[X = x_k] = P_X(x_k)$

$$= F_X(x_k) - F_X(x_{k-1})$$

So the inverse mapping is

$$F_X(x_{k-1}) < y < F_X(x_k) \Rightarrow x_k = F_X^{-1}(y)$$

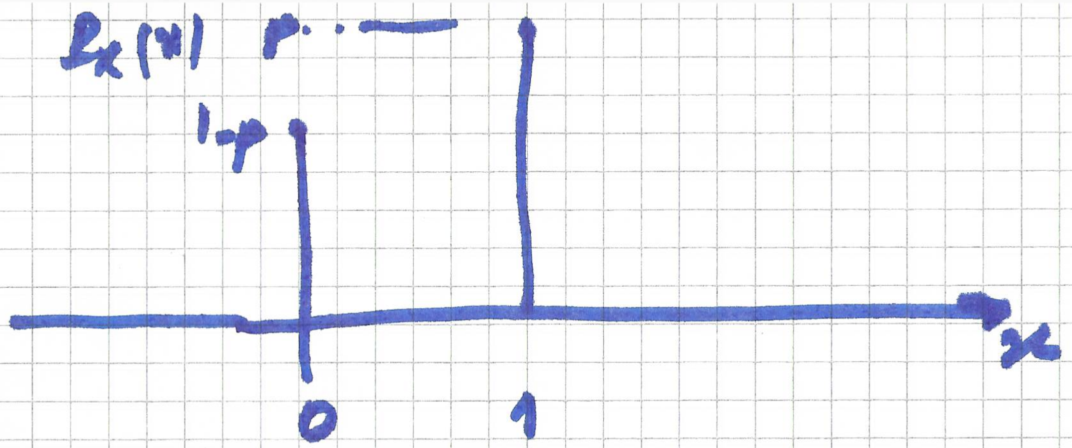
So the algorithmic steps for the generation of a discrete RV is:

1. Generate a random number $Y \sim \mathcal{U}[0,1]$
   Call it y.

2. Compute the value x_k such that
$$F_X(x_{k-1}) \leq y < F_X(x_k)$$

3. take x_k to be the random number generated.

For instance for the generation of a Bernoulli RV $X \sim \text{Bernoulli}(p)$



$$\Pr[X=0] = 1-p$$

$$\Pr[X=1] = p$$

Generate y from $U[0,1]$ and then set

$$x = \begin{cases} 0 & \text{for } y \leq 1-p \\ 1 & \text{otherwise } 1 \geq y > 1-p \end{cases}$$

4.4 Method of Acceptance-Rejection

→ Inverting the CDF of a desired RV X to be simulated can not always be done in closed-form.

⇓

Makes it computationally costly to use this inverse transform approach to generate many samples.

→ Instead one can rely on:

Acceptance-Rejection Approach.

- Suppose that we want to generate a RV Y with PDF $f_Y(y)$.

- To use the acceptance-rejection method, we have to find another RV X with PDF $f_X(x)$ and a constant $c \geq 0$ such as

$$\frac{f_X(x)}{f_Y(y)} \geq c > 0 \quad \text{for } \forall x \in R_X.$$

- Given that the steps of this method are as follow:

Step 1

Generate number x with the PDF $f_X(x)$

Step 2

Generate a random number U (independent of x) from $U \sim U[0,1]$

Step 3

if $U \leq \dfrac{c\, f_Y(x)}{f_X(x)}$ then the output $Y = x$

Else go to step 1.

↳ Proof (see Hwk #5 ib # 3)

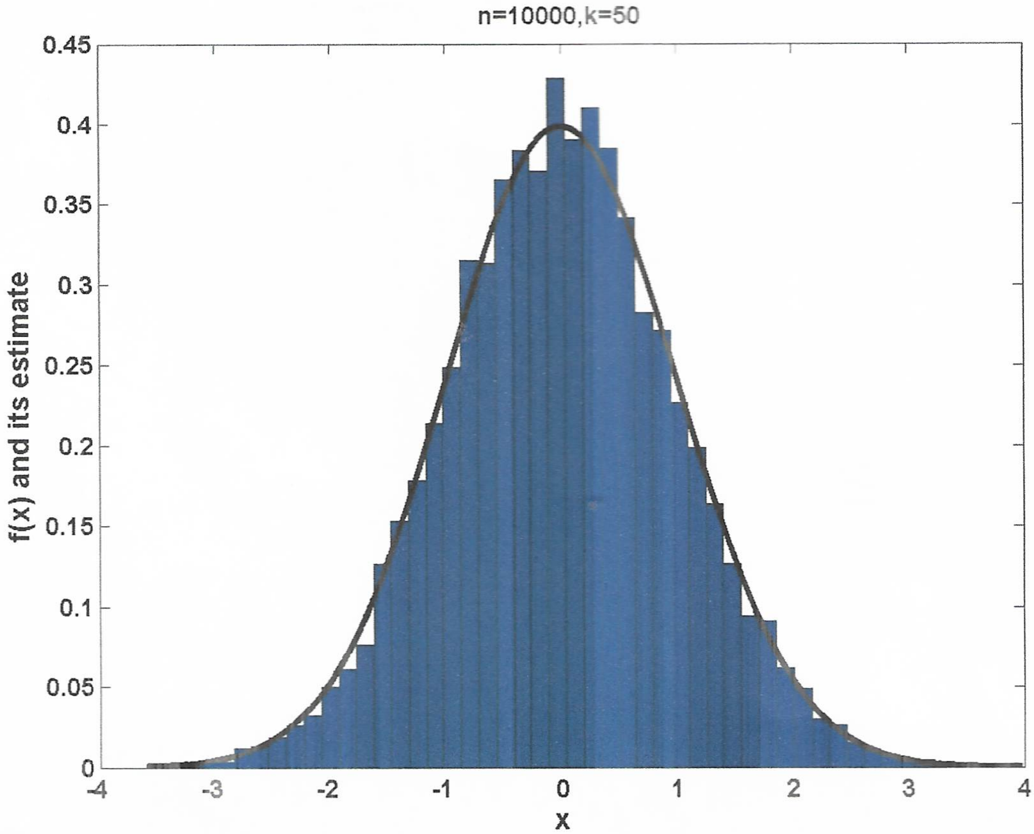
Remark:

For this algorithm to be efficient, c should be as close as possible to 1.

15

4.5 Histogram and Empirical CDF

Assume that we have n i.i.d. observations $\{x_k\}_{k=1}^{n}$ and we want to estimate the underlying PDF. One simple method for doing this is to divide the range of observations into k intervals at points $c_0, c_1, c_2, \dots, c_k$. Defining $\Delta_j = c_j - c_{j-1}, j = 1, 2, \dots, k$ and n_j as the number of observations that fall in the interval $(c_{j-1}, c_j]$, the estimated PDF $\hat{f}(x) = \frac{n_j}{n\Delta_j}, x \in (c_{j-1}, c_j]$. (Function $h(x) = n_j$ is called **histogram**.)



Note that $n_j = \sum_{k=1}^{n} \mathbb{I}\{x_k \in (c_{j-1}, c_j]\}$and $\mathbb{E}\{\hat{f}(x)\} = \frac{\mathbb{E}\{n_j\}}{n\Delta_j} = \frac{n\mathbb{P}(x_k \in (c_{j-1}, c_j])}{n\Delta_j} = \frac{F_X(c_j) - F_X(c_{j-1})}{\Delta_j}$

$$\mathrm{Var}\{\hat{f}(x)\} = \frac{\mathrm{Var}\{n_j\}}{n^2\Delta_j^2} = \frac{n\mathbb{P}(x_k \in (c_{j-1}, c_j])[1 - \mathbb{P}(x_k \in (c_{j-1}, c_j])]}{n^2\Delta_j^2}$$

$$= \frac{[F_X(c_j) - F_X(c_{j-1})][1 - [F_X(c_j) - F_X(c_{j-1})]]}{n\Delta_j^2}$$

For sufficiently small Δ_j,$F_X(c_j) - F_X(c_{j-1}) \approx f_X(c_j)\Delta_j$.

16·1

The empirical CDF (or cumulative histogram) is $\widehat{F}(x)=\frac{1}{n}\sum_{k=1}^{n}\mathbb{I}\{x_k \le x\}$. The sum $\sum_{k=1}^{n}\mathbb{I}\{x_k \le x\}$ is equal to the number of observations less than or equal to x.

$$\mathbb{E}\{\widehat{F}(x)\}=\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\{\mathbb{I}\{x_k \le x\}\}=\frac{1}{n}\sum_{k=1}^{n}\mathbb{P}(x_k \le x)=\frac{1}{n}\sum_{k=1}^{n}F_X(x)=F_X(x)$$

$$\mathrm{Var}\{\widehat{F}(x)\}=\frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}\{\mathbb{I}\{x_k \le x\}\}=\frac{1}{n^2}\sum_{k=1}^{n}\left[\mathbb{P}(x_k \le x)-\left(\mathbb{P}(x_k \le x)\right)^2\right]$$
$$=\frac{F_X(x)-[F_X(x)]^2}{n}\le\frac{1}{4n}$$