Systems Engineering 520

Generating Samples of Random Variables

Huseyin Topaloglu
School of Operations Research
and Industrial Engineering

Lecture 4

Cornell University

Overview

- Generating samples from a discrete distribution
- Inversion method to generate samples from a continuous distribution
- Acceptance / rejection method to generate samples from a continuous distribution
- Generating samples from the normal distribution

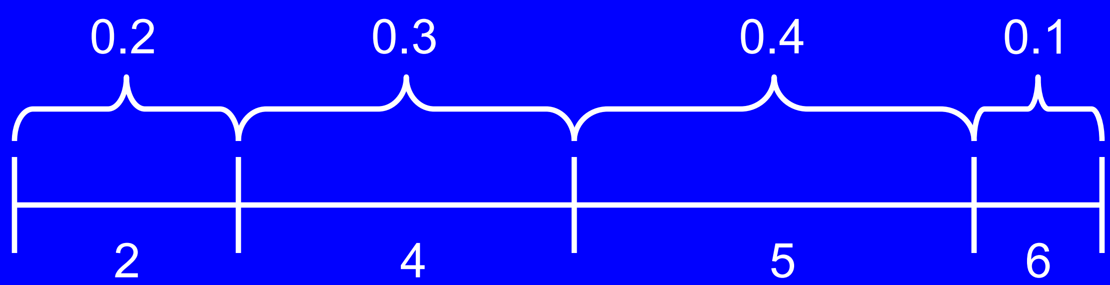
Generating Samples

- We assume that we can always generate a sample from the uniform distribution over [0,1]
- In Excel, RAND() function does this
- In VBA, RND() function does this
- The problem is how to take the sample from the uniform distribution over [0,1] and convert it into a sample from an arbitrary distribution
- E.g. remember that =FLOOR(RAND()*10,1)+1 function in Excel generates an integer uniformly distributed over {1,2,…,10}

Generating Samples from a Discrete Distribution

- Assume that the random variable X has the following p.m.f. and we want to generate samples of X

$$p(x) = \begin{cases} 0.2 & \text{if } x = 2 \\ 0.3 & \text{if } x = 4 \\ 0.4 & \text{if } x = 5 \\ 0.1 & \text{if } x = 6 \end{cases}$$



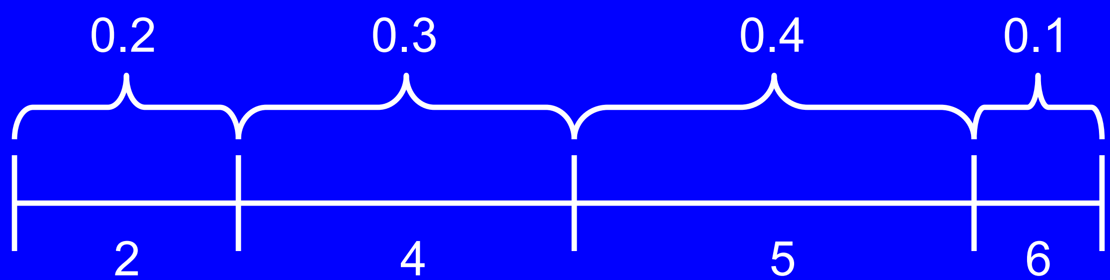
Generating Samples from a Discrete Distribution

- Generate $U \sim U[0,1]$
- If $U < 0.2$, then return $X = 2$

  If $0.2 < U < 0.5$, then return $X = 4$

  If $0.5 < U < 0.9$, then return $X = 5$

  If $U > 0.9$ then return $X = 6$



Generating Samples from a Discrete Distribution

- Let x_1, x_2, x_3,… be the possible values that the random variable X can take
- x_1, x_2, x_3,… can be in any order, not necessarily in increasing order
- The algorithm on the previous slide can alternatively be written as follows
- Generate U~U[0,1]
- Find i such that

$$p(x_1) + \ldots + p(x_{i-1}) < U < p(x_1) + \ldots + p(x_i)$$

- Return X = x_i

Generating Samples from Geometric Distribution

- Let X ~ Geometric (p)
- The possible values that X can take are 1, 2, 3,…
- p.m.f. of X is $p(k) = (1-p)^{k-1}\,p$

$$p(1) + \ldots + p(i) = \sum_{k=1}^{i} (1-p)^{k-1}\,p = 1 - (1-p)^i$$

Generating Samples from Geometric Distribution

- We can generate samples from geometric distribution with parameter p by the following algorithm
- Generate U ~ U[0,1]
- Find i such that

$$1 - (1-p)^{i-1} < U < 1 - (1-p)^i$$

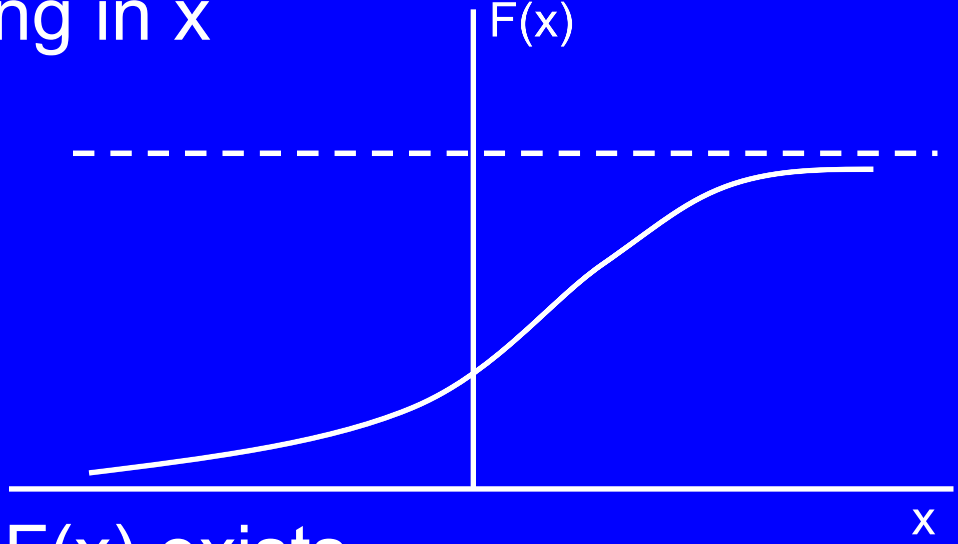
- Return X = i
- For further simplifications, see Page 53 in Ross

Inversion Method to Generate Samples from a Continuous Distribution

- Let X be a continuous random variable with c.d.f. F(x)

$$F(x) = \mathbb{P}\{X \leq x\}$$

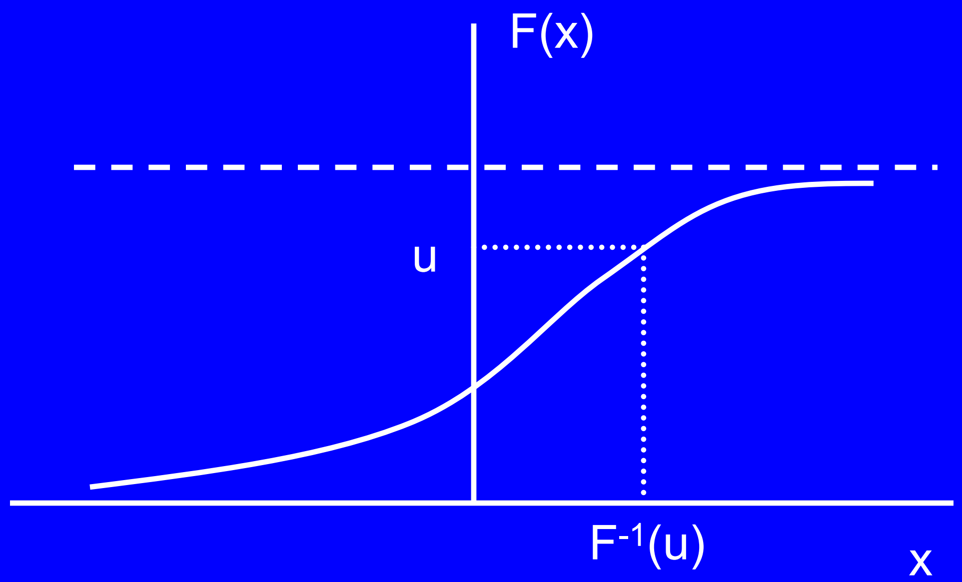
- F(x) must be increasing in x



- Therefore, inverse of F(x) exists

Inversion Method to Generate Samples from a Continuous Distribution

- Recall that $$F(x) = u \iff F^{-1}(u) = x$$



Inversion Method to Generate Samples from a Continuous Distribution

- The following algorithm generates samples of the continuous random variable with c.d.f. $F(x)$
- Generate $U \sim U[0,1]$
- Return $X = F^{-1}(U)$

Inversion Method to Generate Samples from a Continuous Distribution

- Assume that X has exponential distribution with parameter λ and we want to generate samples of X
- c.d.f. of X is $F(x) = 1 - e^{-\lambda x}$
- See Page 68 in Ross

Inversion Method to Generate Samples from a Continuous Distribution

- Assume that X has uniform distribution over [a,b] and we want to generate samples of X

Inversion Method to Generate Samples from a Continuous Distribution

- Assume that X has c.d.f

$$F(x) = x^2 \quad 0 \leq x \leq 1$$

  and we want to generate samples of X

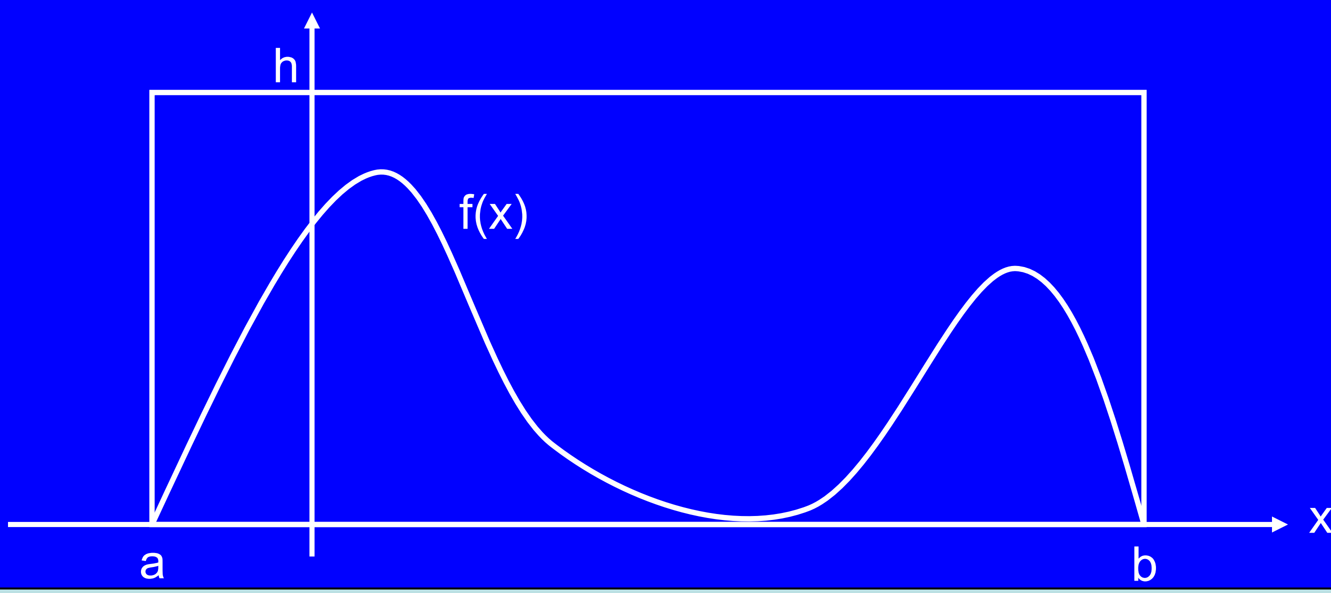
- See Page 68 in Ross

Acceptance / Rejection Method to Generate Samples from a Continuous Distribution

- Inversion method is helpful only when we can explicitly compute the inverse of the c.d.f.
- A/R method can be helpful when the inverse of the c.d.f. cannot be computed explicitly
- However, A/R method works only for random variables that take values over a bounded region
- E.g. A/R method does not work for normal or exponential distribution

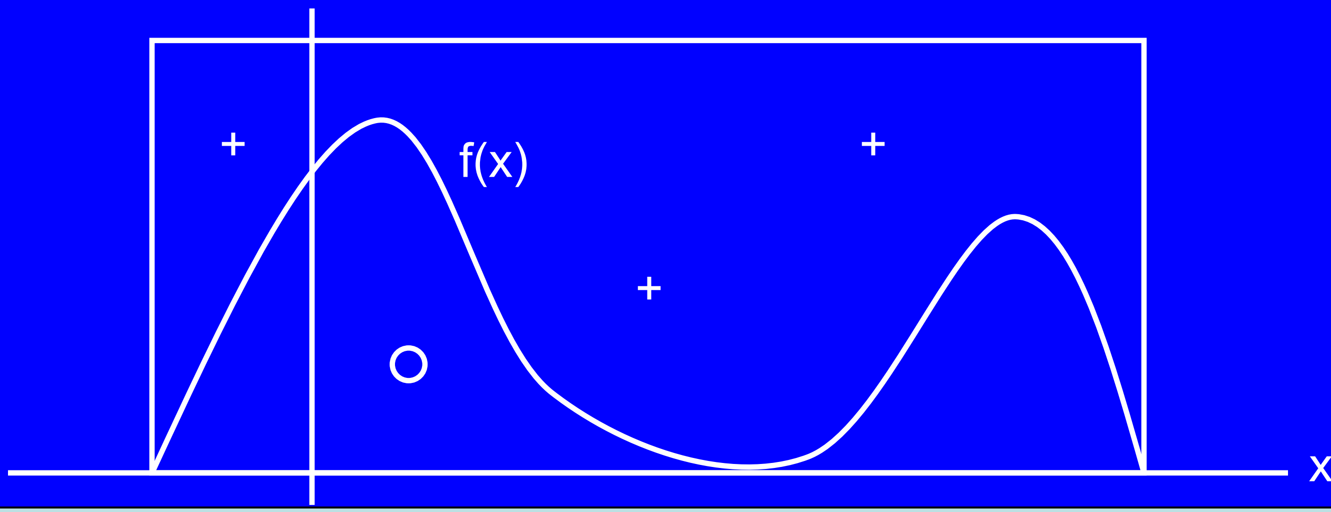
Acceptance / Rejection Method to Generate Samples from a Continuous Distribution

- Assume that we want to generate samples of the random variable X whose p.d.f. is f(x) and f(x) takes positive values only over the interval [a,b]
- Therefore, we can enclose the p.d.f. in a rectangle with width b – a and height h



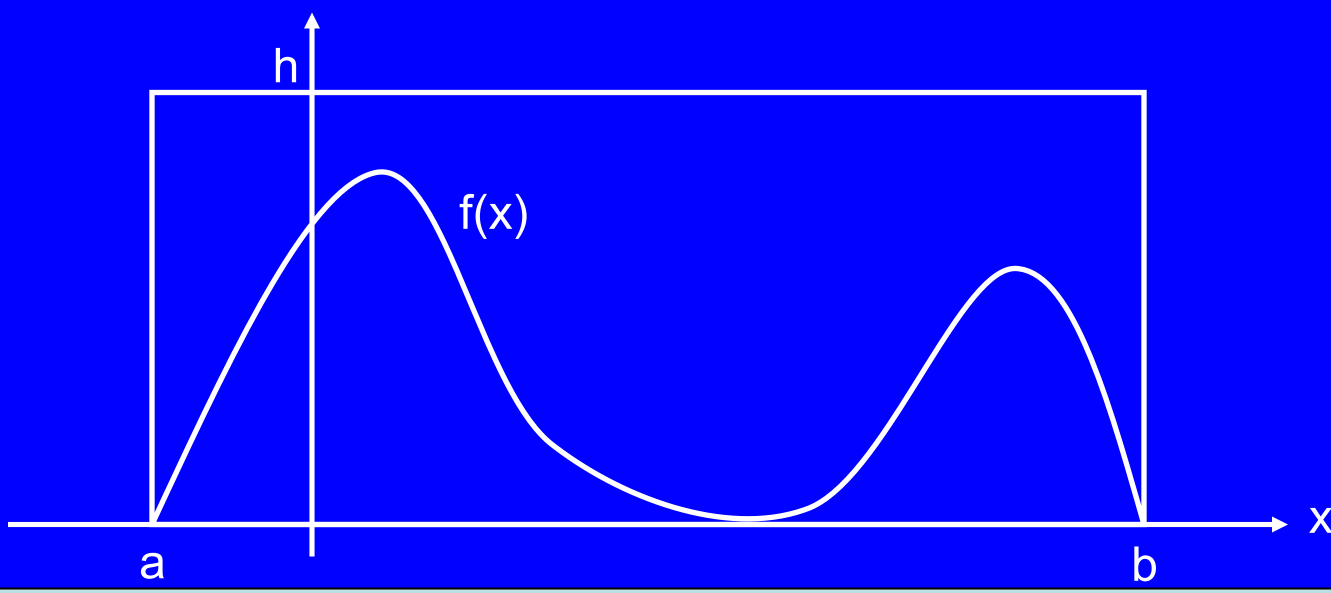
Acceptance / Rejection Method to Generate Samples from a Continuous Distribution

- A/R method works as follows
- Generate a point uniformly distributed over the rectangle
- If the point falls under the p.d.f., then return the x-coordinate of the point as the sample
- Repeat until we get a point under the p.d.f.



Acceptance / Rejection Method to Generate Samples from a Continuous Distribution

- Generate $U_1 \sim U[0,1]$ and $U_2 \sim U[0,1]$
- Set $XC = a + (b - a) U_1$ and $YC = h U_2$
- If $YC < f(XC)$, then return $X = XC$ and stop
- Otherwise, go back to the beginning

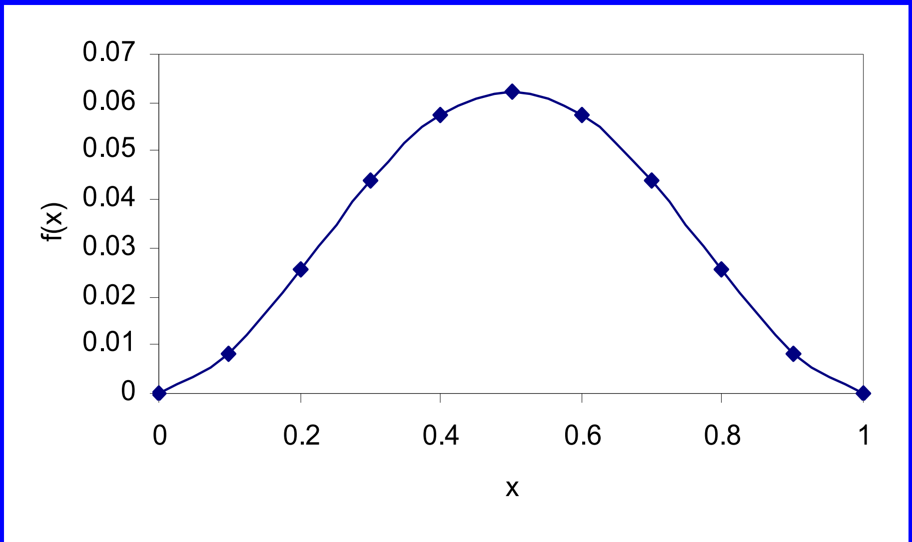


Acceptance / Rejection Method to Generate Samples from a Continuous Distribution

- Assume that X has p.d.f

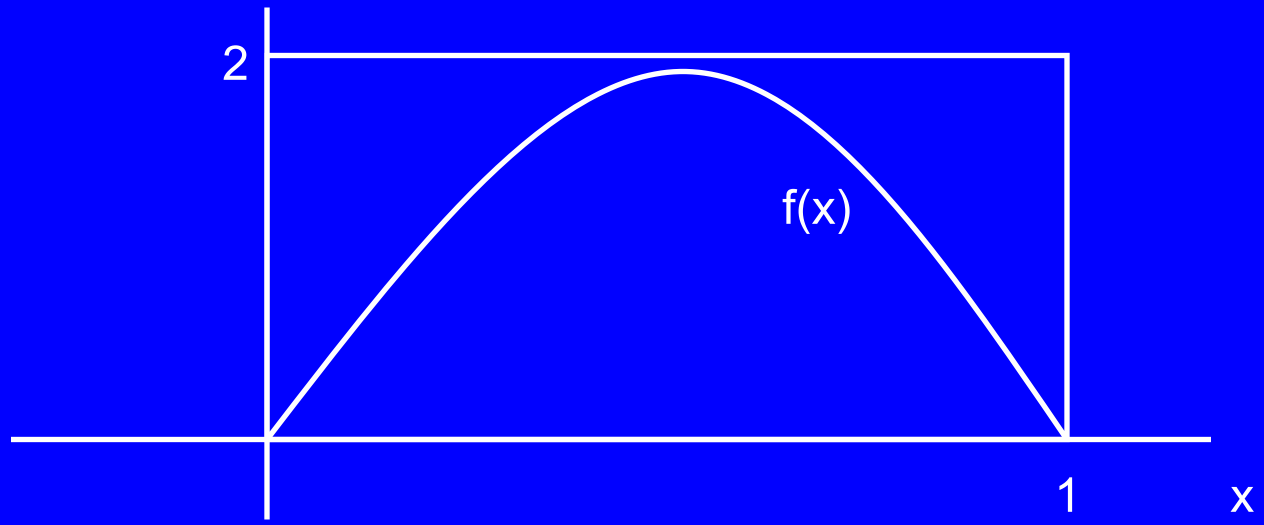
$$f(x) = 30\,(x^2 - 2\,x^3 + x^4) \quad 0 \le x \le 1$$

and we want to generate samples of X



Acceptance / Rejection Method to Generate Samples from a Continuous Distribution

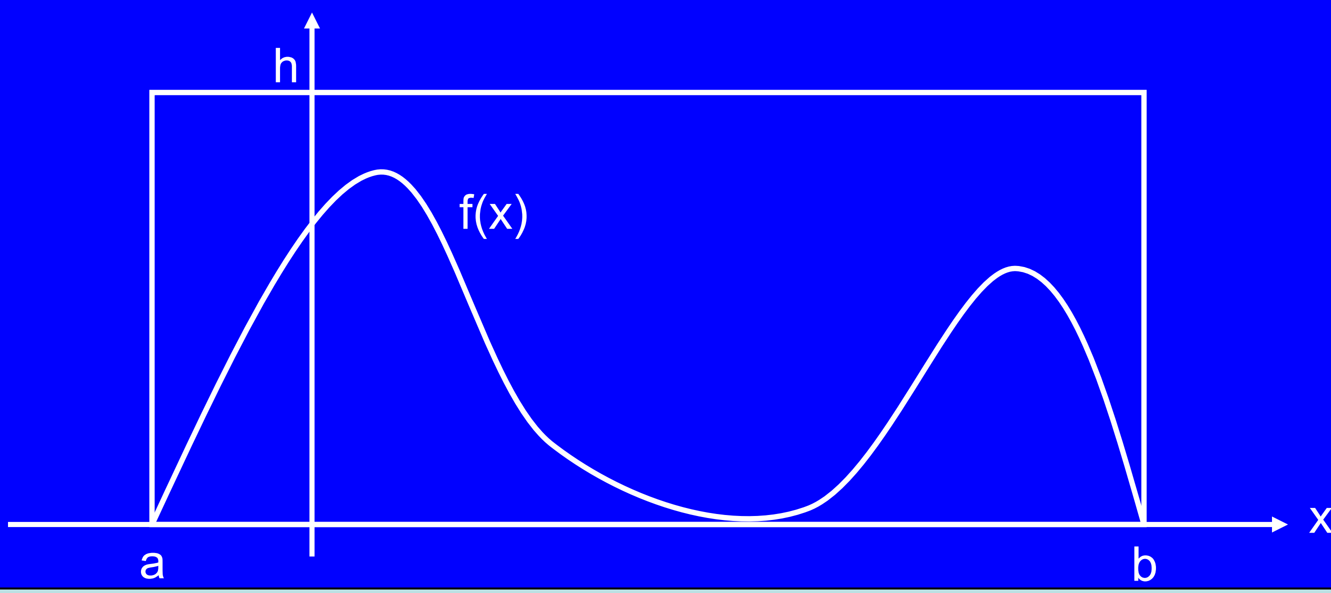
- We can enclose the p.d.f. in the rectangle [0,1] x [0,2]



- What is the probability that A/R will stop after 1 trial?
- What is the probability that A/R will stop after 2 trials?
- What is the probability that A/R will stop after 3 trials?

Acceptance / Rejection Method to Generate Samples from a Continuous Distribution

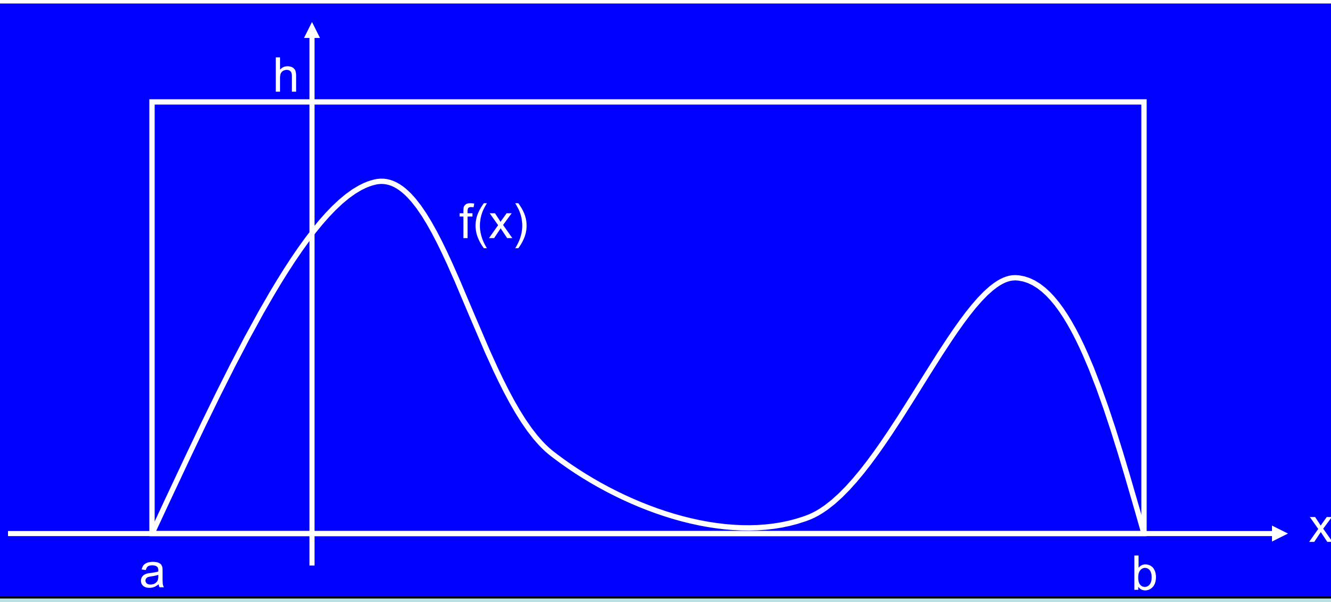
- For the general case, what is the probability that A/R will stop after 1 trial?
- For the general case, what is the probability that A/R will stop after k trials?



Acceptance / Rejection Method to Generate Samples from a Continuous Distribution

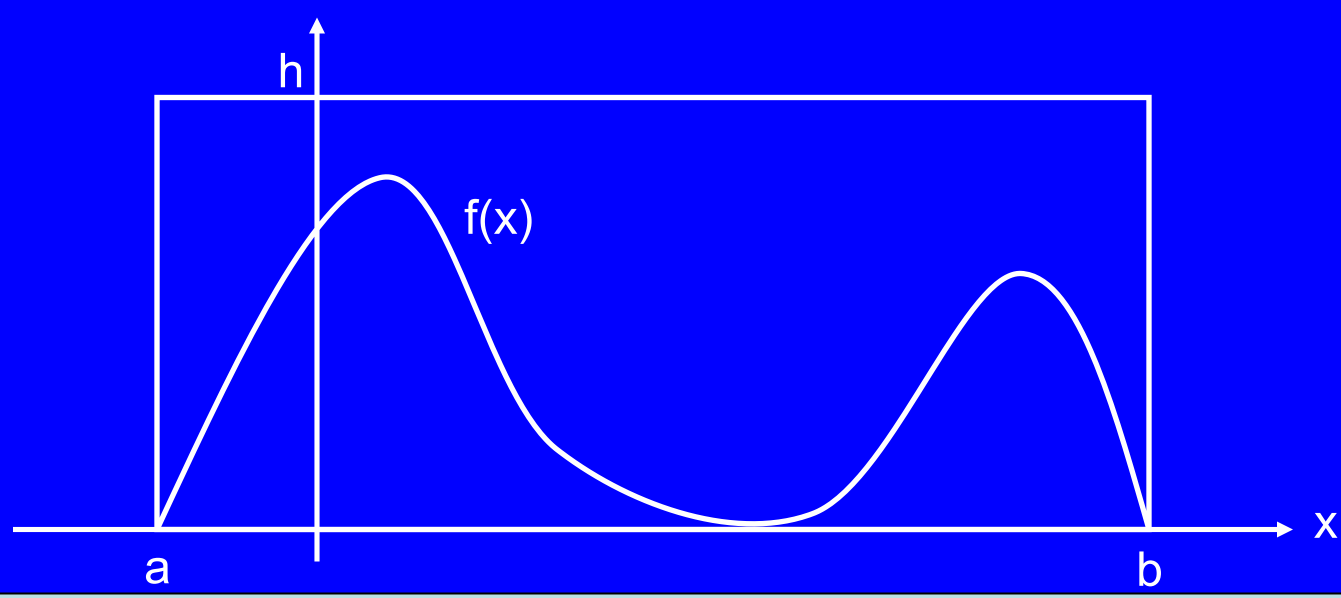
- Let N be the number of trials for A/R to generate one sample of the random variable we are interested in

$$\mathbb{P}\{N = k\} = \left[1 - \frac{1}{h(b-a)}\right]^{k-1} \frac{1}{h(b-a)}$$



Acceptance / Rejection Method to Generate Samples from a Continuous Distribution

- What is the distribution of N?
- We have, $\mathbb{E}\{N\} = h(b-a)$
- Therefore, we should find the "tightest possible" rectangle to make A/R as efficient as possible



Generating Samples from N(0,1)

- The following algorithm generates samples from N(0,1)
- Generate $U_1 \sim U[0,1]$ and $U_2 \sim U[0,1]$
- Set
$$R = \sqrt{-2 \ln U_1}$$
$$\Theta = 2\pi U_2$$
- Set
$$X = R \cos \Theta$$
$$Y = R \sin \Theta$$
- X and Y are independent and both are N(0,1)

Generating Samples from N(0,1)

- This algorithm uses two samples from the uniform distribution over [0,1] to generate two samples from N(0,1)
- If we need a sample from N(μ,σ^2), then we can use

$$\mu + \sigma X \quad \text{and} \quad \mu + \sigma Y$$