Stochastic Simulation
Generation of random variables
Discrete sample space

Bo Friis Nielsen

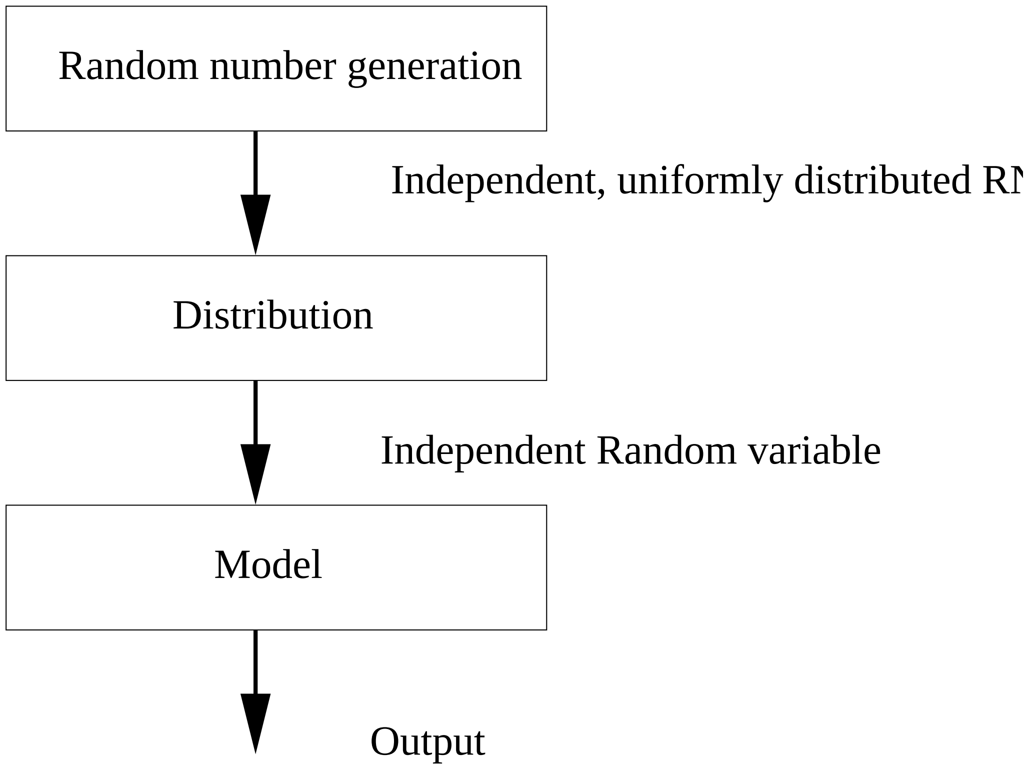
Applied Mathematics and Computer Science

Technical University of Denmark

2800 Kgs. Lyngby – Denmark

Email: bfn@imm.dtu.dk

Plan W1.1-2


Random number generation
Independent, uniformly distributed RN
Distribution
Independent Random variable
Model
Output

Random variables

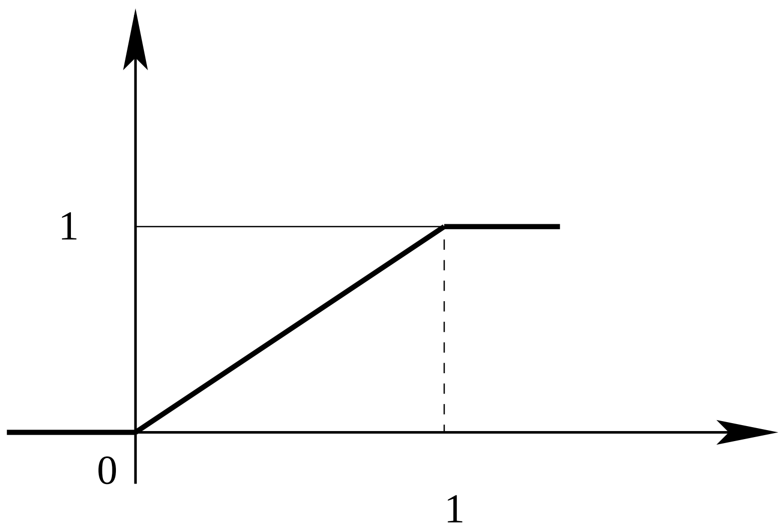
Aim

- The scope is the generation of **independent** random variables X_1, X_2, ... X_n with a **given distribution**, $F_x(x)$, (or probability density function [pdf]).
- We assume we have access to a supply (U_i) of random numbers, independent samples from the uniform distribution on $]0,\ 1[$.
- Our task is to transform U_i into X_i.

Uniform distribution I

Our norm distribution or building block, $U(0,1)$

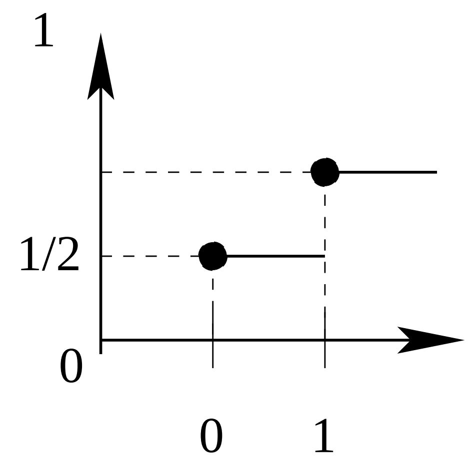
$$f(x) = 1 \quad F(x) = x \quad \text{for} \quad 0 \le x \le 1$$

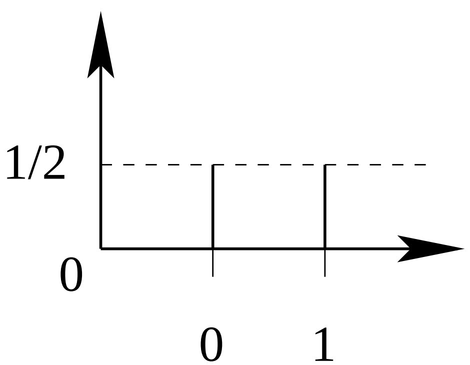


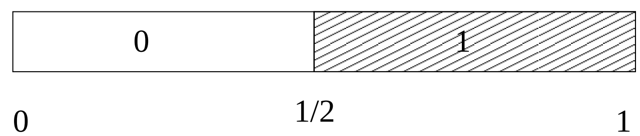
$$\mathsf{E}(X) = \frac{1}{2} \quad \mathsf{Var}(X) = \frac{1}{12}$$

Coin

or uniform distribution







$$X = 0,\ 1$$

$$\mathsf{P}(X = i) = \frac{1}{2}$$

$$X := \left(U > \frac{1}{2}\right) \qquad X = \lfloor (2U) \rfloor$$

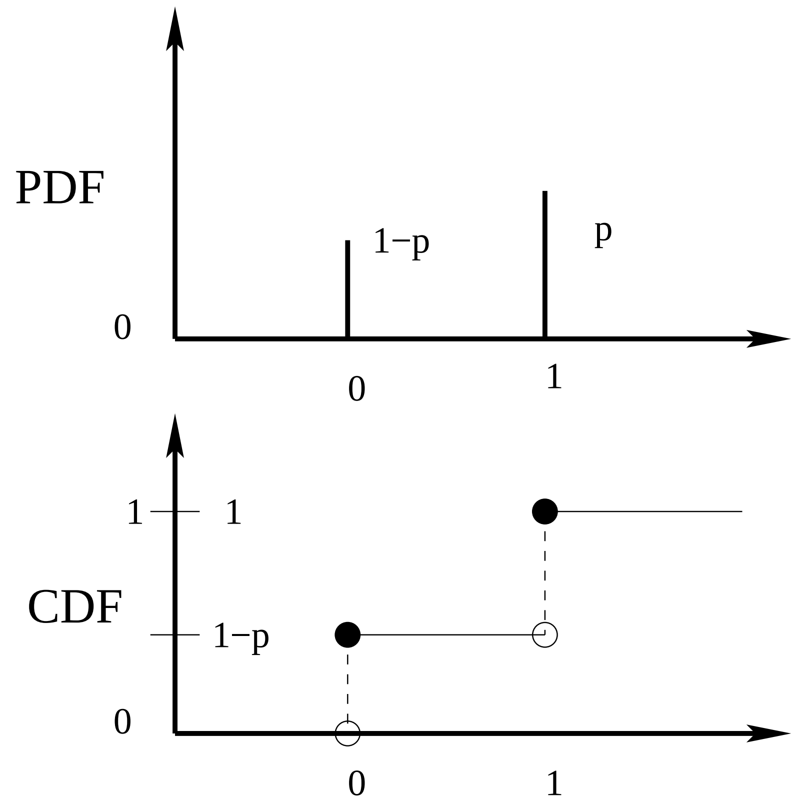
Bernoulli trial

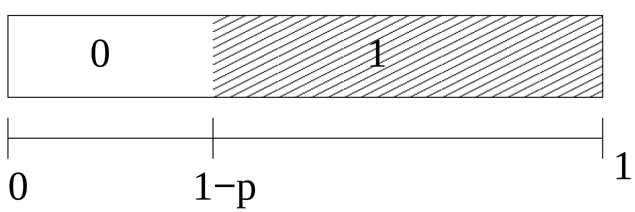
Toss a coin with $\mathsf{P}(X = 1) = p$ and $\mathsf{P}(X = 0) = 1 - p$.

$$F(x) = \mathsf{P}\left(X \leq x\right)$$

$$X = U > 1 - p$$

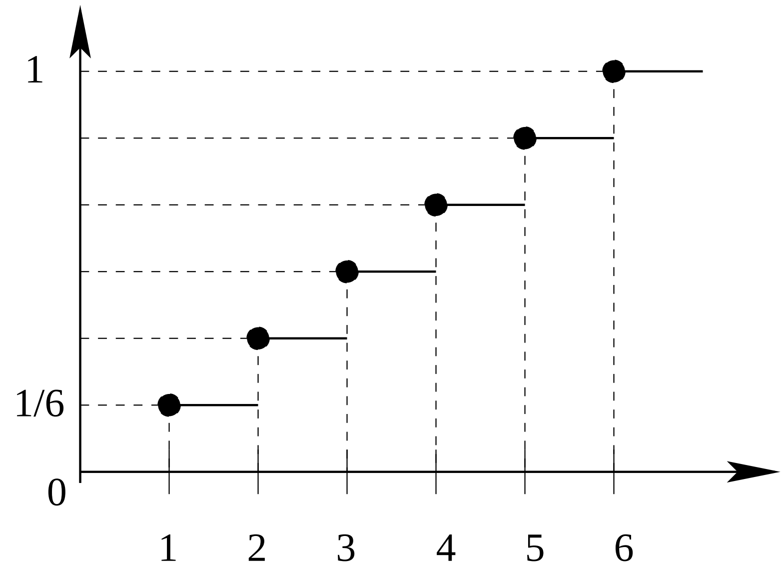
$$\begin{aligned} X &= 0 \qquad 0 \leq U \leq 1 - p \\ X &= 1 \qquad 1 - p < U \leq 1 \end{aligned}$$

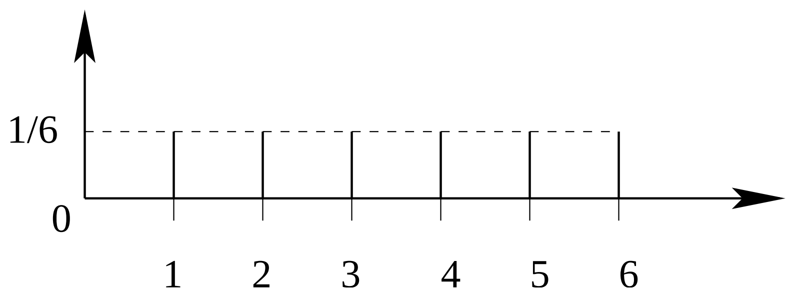


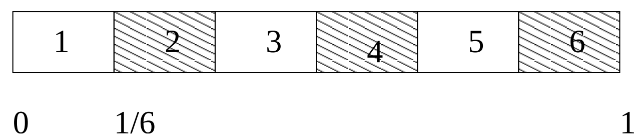


A fair die

or uniform distribution







$$X = 1,\ 2,\ \dots 6$$

$$\mathsf{P}(X = i) = 1/6$$

$$X = \lfloor (6U) \rfloor + 1$$

Can be generalized $6 \rightarrow k$.

Discrete distribution - direct (crude) method

Suppose X can take k distinct values $x_1 < x_2 < \ldots x_k$ with

$$p_i = \mathsf{P}(X = x_i), \quad i = 1,\ 2,\ \ldots\ ,\ k$$

Then X takes the value x_i with probability p_i if U falls in an interval with length p_i.
That is if

$$\sum_{j=1}^{i-1} p_j < U \leq \sum_{j=1}^{i} p_j$$

or

$$\boxed{X = x_i \quad if \quad F(x_{i-1}) < U \leq F(x_i)}$$

Geometric distribution, $NB(1,p)$

The discrete time version of waiting time. Memory-less.

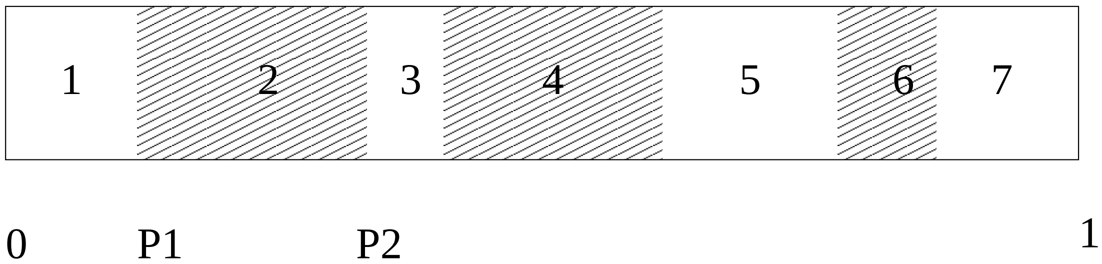
$$
\begin{aligned}
f(n) &= \mathsf{P}(X=n) = (1-p)^{n-1}p \quad n = 1,\ 2,\ \ldots \\
F(n) &= \mathsf{P}(X \leq n) = 1-(1-p)^n
\end{aligned}
$$

$$X = n \quad if \quad F(n-1) < U \leq F(n) \qquad 1-(1-p)^{n-1} < U \leq 1-(1-p)^n$$

$$n - 1 < \frac{\log(1-U)}{\log(1-p)} \leq n$$

$$\boxed{X = \left\lfloor \left(\frac{\log(U)}{\log(1-p)} \right) \right\rfloor + 1}$$

Discrete distribution II



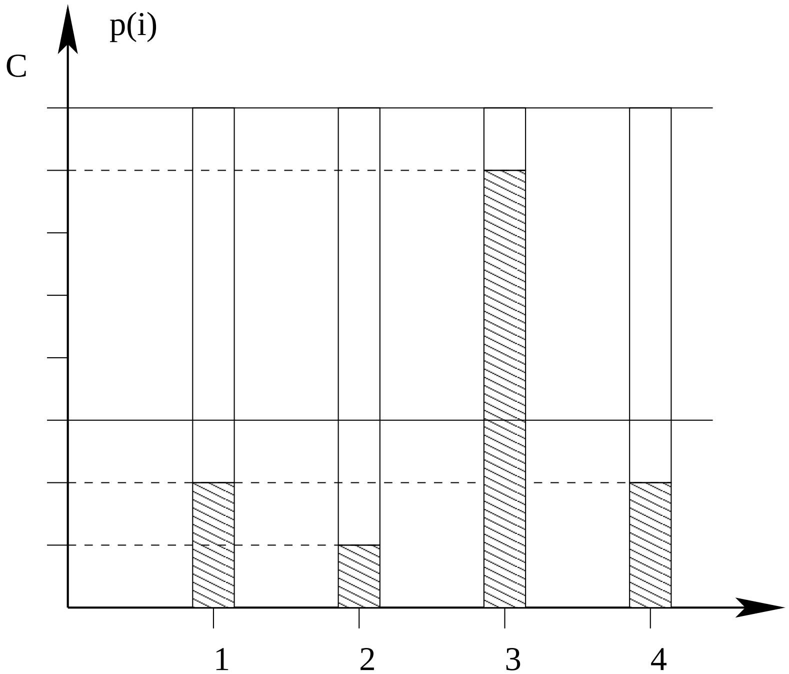
1. Generate U
2. Find the interval i which U belong to. $P_{i-1} < U \leq P_i$
3. output x_i

- Linear search ($\mathrm{E}(X)$)
- Rearrangement of intervals
- Binary search
- Indexed search

Rejection Method

Simple rejection More optimistic: acceptance method.

Assume $\mathsf{P}(X = i) = p_i$ for $i = 1,\ 2,\ \ldots\ k$.



Let $c \geq p_i$ (then $p_i/c \leq 1$).

1. $I = \lfloor (k * U_1) \rfloor + 1$
2. if $U_2 \leq p_I/c$ output: I
   Else goto 1.

frequency for i : $\dfrac{\frac{1}{k}\frac{p_i}{c}}{\sum_{j=1}^{k}\frac{1}{k}\frac{p_j}{c}} = p_i$

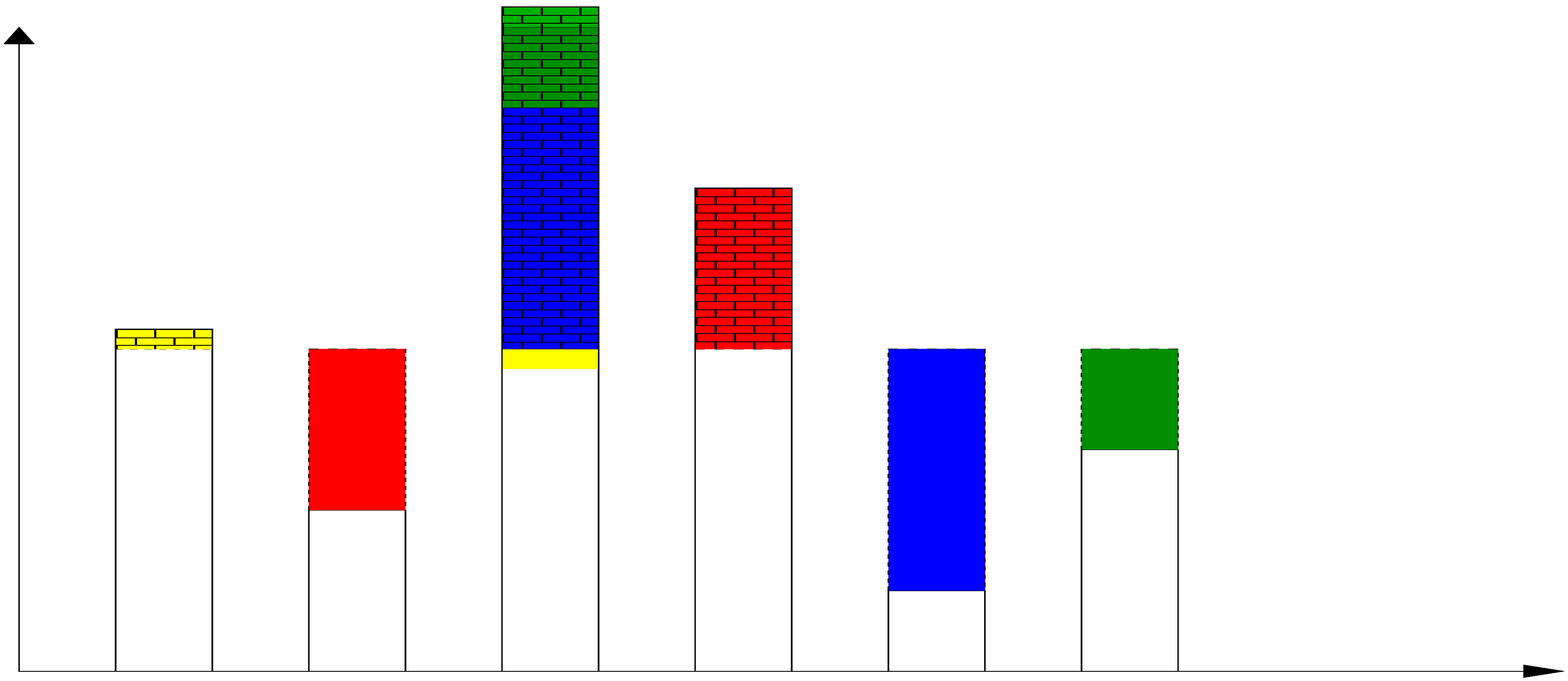
Alias method

- A method for generating discrete random variates of general type
- From discrete uniform to general discrete
- Generate one random number
- One comparison
- Potentially one table look-up
- Drawback: Complex set-up procedure

A six-point distribution

$$\mathsf{P}(X=1)=\frac{17}{96} \qquad \mathsf{P}(X=2)=\frac{1}{12} \qquad \mathsf{P}(X=3)=\frac{1}{3}$$

$$\mathsf{P}(X=4)=\frac{1}{4} \qquad \mathsf{P}(X=5)=\frac{1}{24} \qquad \mathsf{P}(X=6)=\frac{11}{96}$$



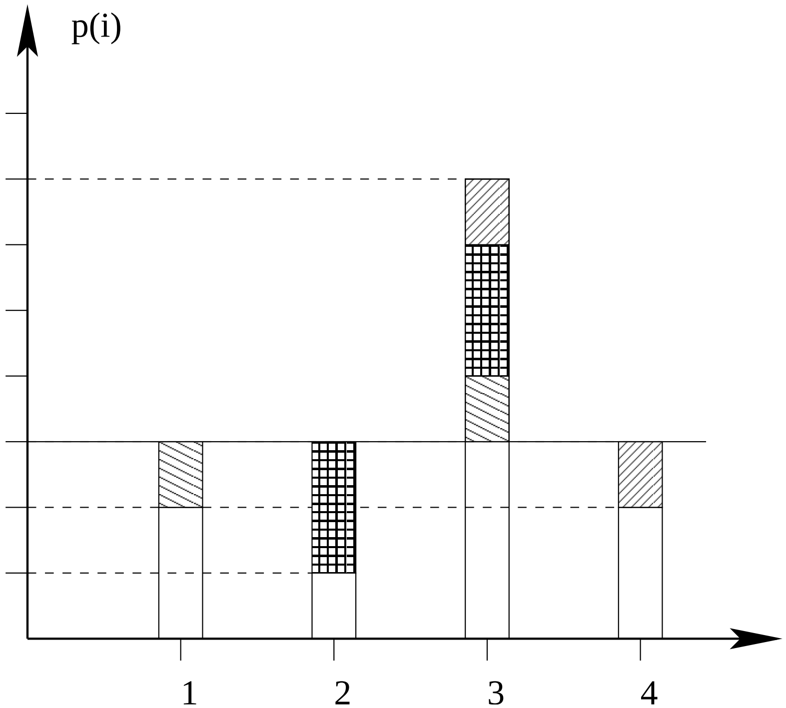
Alias method

- Setup procedure
  - Generate the table of $F(I)$-values, (which part of the mass belongs to I itself.
  - Generate the table of $L(I)$-values, (the alias of class I)
- Method at run time
  - Generate I:$I = \lfloor k * U_1 \rfloor$+1
  - Test against $F(I)$.If $U_2 \leq F(I)$ then return $X = I$ else return $X = L(I)$. The L, F tables for the six-point distribution

$$F(1) = 1 \quad F(2) = \frac{1}{2} \quad F(3) = \frac{15}{16} \quad F(4) = 1 \quad F(5) = \frac{1}{4} \quad F(6) = \frac{11}{16}$$

$$L(1) = 1 \quad L(2) = 4 \quad L(3) = 1 \quad L(4) = 4 \quad L(5) = 3 \quad L(6) = 3$$

Alias Method



On setup: generate F and L.

Generation:

1. $I = \lfloor (k * U_1) \rfloor + 1$
2. if $U_2 \leq F(I)$ output I
   else output $L(I)$.

The Alias tables

Generate F and L.

Pseudo code. p is a vector containing the probabilities.

1. `L=`$\{1, \ldots, k\}$
2. `F=k*p` ($F = 1$ is equivalent for the uniform dist.)
3. `G=find(F>=1)` and `S=find(F<=1)`
4. `while ~isempty(S),`
   (a) `i=G(1)` and `j=S(1)`
   (b) `L(j)=i` and `F(i)=F(i)-(1-F(j))`
   (c) `if F(i)<1-eps` then `G(1)=[]` and `S=[S i]`
   (d) `S(1)=[]`

Rejection Method

General method

Aim: We will generate X with probabilities $p_i = \mathsf{P}(X = i)$.

Assume Y with probabilities $q_i = \mathsf{P}(Y = i)$ is easily generated and $C \geq \frac{p_i}{q_i}$ for all $i = 1, \ldots$.

1. Generate Y with probability q_i and let $X^\star = Y$.
2. Generate U_2.
   If $U_2 \leq \frac{p_{X^\star}}{Cq_{X^\star}}$ output $X = X^\star$
   else goto 1.

Rejection Method: Probability for $X = i$:

$$
\begin{aligned}
\mathsf{P}(X=i) &= \mathsf{P}(X^\star = i|\mathsf{accept}) \\
&= \frac{\mathsf{P}(X^\star = i, \mathsf{accept})}{\mathsf{P}(\mathsf{accept})} \\
&= \frac{\mathsf{P}(X^\star = i)\mathsf{P}(\mathsf{accept})|X^\star = i)}{\mathsf{P}(\mathsf{accept})} \\
&= \frac{q_i \cdot \frac{p_i}{Cq_i}}{\sum_j q_j \frac{p_j}{Cq_j}} \\
&= p_i
\end{aligned}
$$

Excercise 2

Discrete random variables

In the excercise you can use a build in procedure for generating random numbers. Compare the results obtained in simulations with expected results. Use histograms (and tests).

1. Choose a value for the probability parameter p in the geometric distribution and simulate 10,000 outcomes. You can experiment with a small, moderate and large value if you like.
2. Simulate the 6 point distribution with

X	1	2	3	4	5	6
p_i	$7/48$	$5/48$	$1/8$	$1/16$	$1/4$	$5/16$

   (a) by applying a direct (crude) method

(b) by using the the rejction method
(c) by using the Alias method

3. Compare the three different methods using adequate criteria, then discuss the results.

4. Give recommendations of how to choose the best suited method in different settings, i.e., discuss the advantages and drawbacks of each method. If time permits substantiate by running experiments.