Generation from simple discrete distributions

- Note! This is just a more clear and readable version of the same slide that was already in the Generation of Random Numbers, Part 1 (slide 12).
- In the following $U,\ U_1,\dots,\ U_n$ denote independent random variables ~U(0,1)
- $\mathrm{int}(X) = \lfloor X \rfloor$ = integer part of X

Distribution	Expression for generation
Symmetric bivalued {0,1} distribution $\mathrm{P}\{X=0\} = \mathrm{P}\{X=1\} = 0.5$	$\mathrm{int}(2U)$ or $\mathrm{int}(U+0.5)$
Symmetric bivalued {0,1} distribution $\mathrm{P}\{X=0\} = 1-p,\ \mathrm{P}\{X=1\} = p$	$\mathrm{int}(U+p)$
Bivalued {-1,1} distribution $\mathrm{P}\{X=0\} = 1-p,\ \mathrm{P}\{X=1\} = p$	$2\ \mathrm{int}(U+p)-1$
Trivalued {0,1,2} distribution probs:t $= \{1-p_1-p_2,\ p_1,\ p_2\}$	$\mathrm{int}(U+p_2)+\mathrm{int}(U+p_1+p_2)$
Uniform discrete distribution $\{0,\ 1,\ 2,\dots,n-1\}$	$\mathrm{int}(n\,U)$
Uniform discrete distribution $\{1,\ 2,\ 3,\dots,n\}$	$\mathrm{int}(n\,U)+1$
Binomial distribution $\mathrm{Bin}(n,p)$	$\sum_{i=1}^{n} \mathrm{int}(U_i+p)$

Generation from geometric distribution

- The point probabilities of a discrete random variable X obeying the geometric distribution $\mathrm{Geom}(p)$ are
$$\mathrm{P}\{X = i\} = p_i = p(1-p)^i \qquad i = 0, 1, 2, \ldots$$
- The generation of samples of X can be done with the following simple procedure
- Algorithm
$$X = \left\lfloor \frac{\log U}{\log(1-p)} \right\rfloor$$
where $U \sim \mathrm{U}(0,1)$
- In fact, this represents generation of samples from the distribution $\mathrm{Exp}(-\log(1-p))$ and discretization to the closest integer smaller then or equal to that value

Rejection method (rejection-acceptance method)

- The task is to generate samples of the random variable X from a distribution with pdf $f(x)$
- Let $g(x)$ be another density function and c a constant such that
  - $c\,g(x)$ majorizes $f(x)$, i.e. $c\,g(x) \geq f(x)$ in the whole range of X
  - there is an (easy) way to generate samples for a random variable with pdf $g(x)$
- The generation of X can be done with the following method:
- Algorithm
  - Generate X with pdf $g(x)$
  - Generate Y from the uniform distribution $\mathrm{U}(0, c\,g(X))$
  - If $Y \leq f(X)$ then accept X
    - otherwise generate as above new values X and Y until a pair is found which satisfies the acceptance criterion; return X
- Proof:

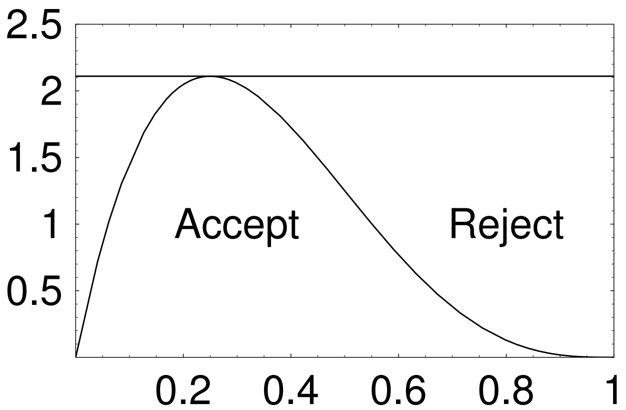
$$\mathrm{P}\{X \in (x, x+dx) \text{ and } Y \leq f(X)\} = g(x)dx \cdot f(x)/cg(x) = f(x)dx/c$$

$$\mathrm{P}\{Y \leq f(X)\} = \int f(x)dx/c = 1/c$$

$$\mathrm{P}\{X \in (x, x+dx)\,|\,Y \leq f(X)\} = (f(x)dx/c)/(1/c) = f(x)dx$$

Rejection method (example)

- When the range is a finite interval (a, b) one can choose $g(x)$ to be the pdf of a random variable uniformly distributed in this interval: $g(x) = 1/(b-a)$, when $x \in (a, b)$



- Assume $X \in (0, 1)$ obeys the beta distribution $\beta(2, 4)$ with pdf

$$f(x) = 20x(1-x)^3, \qquad 0 \le x < 1$$

- The function is limited in a rectangle with height 2.11
  - choose $c = 2.11$ and $g(x) = 1$, when $0 \le x < 1$

- The algorithm is now the following
  - Generate X from the uniform distribution $\mathrm{U}(0, 1)$
  - Generate Y from the uniform distribution $\mathrm{U}(0, 2.11)$
  - If $Y \le 20X(1-X)^3$, accept X and stop, otherwise continue from the beginning until an acceptable pair has been found

- Here the generated values (X, Y) represent a point uniformly distributed in the rectangle
  - it is clear that the proportion of accepted values of $X = x$ is proportional to $f(x)$
  - the pdf of the accepted values X is then $f(x)$

Composition method

- Assume that the pdf $f(x)$ of X, from which samples are to be drawn, can be written (decomposed) in the form

$$f(x) = \sum_{i=1}^{r} p_i f_i(x)$$

  where

  - the p_i form a discrete probability distribution, $\sum_i p_i = 1$
  - the $f_i(x)$ are density functions, $\int f_i(x)\,dx = 1$

- This kind of distribution is called a composition distribution
- The sample generation can be done as follows
  - draw index I from the distribution $\{p_1, p_2, \ldots, p_r\}$
  - draw value of X using the pdf $f_I(x)$

Composition method (continued)

- For instance, the method can be used by dividing the range of X (a, b) into smaller intervals $(a_1, b_1),\ (a_2, b_2), \ldots,\ (a_n, b_n)$
  - p_i is then the probability that X lies in the interval i

$$p_i = \int_{a_i}^{b_i} f(x)\,dx$$

  - $f_i(x)$ is the conditional pdf in the interval i

$$f_i(x) = \begin{cases} f(x)/p_i & x \in (a_i, b_i) \\ 0 & \text{otherwise} \end{cases}$$

Composition method (example 1)

- The task is to generate samples X from the distribution Exp(1)
- Divide $(0, \infty)$ into intervals $(i, i+1)$, $i = 0, 1, 2, \ldots$
- The probabilities of the intervals

$$p_i = \mathrm{P}\{i \le X < i+1\} = e^{-i} - e^{-(i+1)} = e^{-i}(1 - e^{-1})$$

  constitute a geometric distribution (starts from 0)

- The conditional pdf's are

$$f_i(x) = e^{-(x-i)}/(1 - e^{-1}) \qquad i \le x < i+1$$

  that is, in the interval i, r.v. $(X - i)$ has the pdf $e^{-x}/(1 - e^{-1})$, $0 \le x < 1$

- Algorithm
  - draw I from geometric distribution $p_i = e^{-i}(1 - e^{-1})$, $i = 0, 1, 2, \ldots$
  - draw Y with the pdf $e^{-x}/(1 - e^{-1})$, $0 \le x < 1$ (for instance, using the rejection method)
  - $X = I + Y$
- Advantage: one does not need to compute the logarithm function unlike when using the inverse transform method

Composition method (example 2)

- Instead of the pdf one can as well work with the cdf's in the composition method
- Let the cdf of X be

$$\begin{aligned} F(x) &= 1 - \alpha e^{-\beta_1 x} - (1-\alpha)e^{-\beta_2 x} \\ &= \alpha(1 - e^{-\beta_1 x}) + (1-\alpha)(1 - e^{-\beta_2 x}) \end{aligned}$$

- Algorithm
  - draw the index I: $\mathrm{P}\{I = 1\} = \alpha$, $\mathrm{P}\{I = 2\} = 1 - \alpha$
  - draw the value of X from the distribution $F_I(x)$

$$F_1(x) = 1 - e^{-\beta_1 x} \qquad F_2(x) = 1 - e^{-\beta_2 x}$$

  - or if $I = 1$ then $X = -\frac{1}{\beta_1}\log U$; if $I = 2$ then $X = -\frac{1}{\beta_2}\log U$
- using the inverse transformation method would be rather difficult
  - the inverse cdf function cannot be calculated analytically

Characterization of the distribution

- Many distributions are defined in the form: X is distributed as the sum of n independent random variables, each of them obeying a given distribution “Dist”
- Then X can be generated literally by drawing independently values for n random variables Z_i from distribution “Dist”; then $X = Z_1 + Z_2 + \cdots + Z_n$ obeys the desired distribution
- Examples of this kind of distributions are the binomial distribution, gamma distribution (Erlang’s distribution) and χ^2-distribution

Characterization method (example: binomial distr.)

- The binomial distribution $\mathrm{Bin}(n,p)$ is the distribution obeyed by the sum of n independent Bernoulli(p)-variables

$$X = \sum_{i=1}^{n} B_i, \quad B_i \sim \mathrm{Bernoulli}(p) \quad \Rightarrow \quad X \sim \mathrm{Bin}(n,p)$$

  - Bernoulli(p)-variable takes value 1 with probability p and value 0 with probability $1-p$
  - $B_i = \mathrm{int}(p + U_i) = \lfloor p + U_i \rfloor, \quad U_i \sim \mathrm{U}(0,1)$ (integer part)

- Algorithm

$$X = \sum_{i=1}^{n} \lfloor p + U_i \rfloor, \quad U_i \sim \mathrm{U}(0,1)$$

Characterization method (example: gamma distribution)

- When n is an integer $\Gamma(n, \lambda)$-distribution is the distribution of the sum of n independent random variables obeying the $\mathrm{Exp}(\lambda)$ distribution

$$X = \sum_{i=1}^{n} Y_i, \quad Y_i \sim \mathrm{Exp}(\lambda) \quad \Rightarrow \quad X \sim \Gamma(n, \lambda)$$

- By taking into account how exponentially distributed random variables can be generated we get the following algorithm
- Algorithm

$$X = -\frac{1}{\lambda} \log \prod_{i=1}^{n} U_i, \qquad U_i \sim \mathrm{U}(0,1)$$

- The sum of logarithms has been written as a logarithm of the product
  - this is advantageous as the logarithmic function has to be computed only once

Characterization method (example χ^2-distribution)

- $\chi^2(\nu)$-distribution with ν degrees of freedom (integer) represents the sum of ν independent $\mathrm{N}(0,1)$-distributed random variables

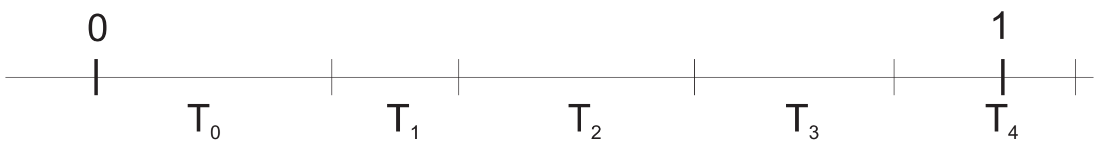
$$X = \sum_{i=1}^{\nu} Y_i, \quad Y_i \sim \mathrm{N}(0,1) \quad \Rightarrow \quad X \sim \chi^2(\nu)$$

Characterization method (example: Poisson distr.)

- Another type of example of the characterization method is provided by the Poisson distribution
- The number of arrivals N from a Poisson process (intensity a) in the interval $(0, 1)$ is Poisson distributed with parameter a, $N \sim \text{Poisson}(a)$
- Draw interarrival times $T_i,\ i = 0, 1, 2, \ldots$ from the Exp(a)-distribution, $T_i = -(1/a)\log U_i$
- N is the number of intervals within interval (0,1) or formally $N = \min\left\{n : \sum_{i=0}^{n} T_i > 1\right\}$
- Algorithm

$$N = \min\left\{n : \prod_{i=0}^{n} U_i < e^{-a}\right\}$$

  - multiply numbers $U_i \sim \text{U}(0, 1),\ i = 0, 1, 2, \ldots$
  - N is the first value of i such that the product is less than e^{-a}



Poisson distribution: numerical example

- Let the mean be $a = 0.2$
- The comparison parameter is $u = e^{-0.2} = 0.8187$

i	U_i	$U_0 \cdots U_i$	accept/continue	Result
0	0.4357	0.4357	$< u$, accept	$N = 0$
0	0.4146	0.4146	$< u$, accept	$N = 0$
0	0.8353	0.8353	$\geq u$, continue	
1	0.9952	0.8313	$\geq u$, continue	
2	0.8004	0.6654	$< u$, accept	$N = 2$

- When a is large the method is slow; the values of N are then typically large and one has to generate a large number of values U_i
- Then it is better to use the discretization method (inversion of the discrete cdf)
- For very large values of a, one may also apply approximation by normal distribution (denote $Z \sim \mathrm{N}(0,1)$)

$$\mathrm{Poisson}(a) \approx \mathrm{N}(a, a) \Rightarrow N \approx \lceil a + \sqrt{a}Z - 0.5 \rceil$$

Characterization method (other examples)

- The a^{th} smallest of the numbers $U_1, U_2, \ldots, U_{a+b+1}$, where the U_i are independent uniformly distributed random variables, $U_i \sim \mathrm{U}(0,1)$, obeys the $\beta(a,b)$-distribution
- The ratio of two $N(0,1)$-distributed random variables obeys the Cauchy$(0,1)$-distribution
- $\chi^2(\nu)$-distribution with an even number of degrees of freedom ν is the same as the $\Gamma(\nu/2, 1/2)$-distribution
- With two independent gamma-distributed random variables one can construct a beta-distributed random variable

$$X_1 \sim \Gamma(b,a) \quad X_2 \sim \Gamma(c,a) \quad \Rightarrow \quad \frac{X_1}{X_1+X_2} \sim \beta(b,c)$$

- If $X \sim N(0,1)$ is a normally distributed random variable, then $e^{\mu+\sigma X}$ is lognormal(μ,σ) random variable

Generation from a multi-dimensional distribution

- Task: generate samples of $X_1, \ldots, X_n$, which have the joint density function $f(x_1, \ldots, x_n)$
- Write this density function in the form $f(x_1, \ldots, x_n) = f_1(x_1)f_2(x_2 \mid x_1) \ldots f_n(x_n \mid x_1, \ldots, x_{n-1})$ where $f_1(x_1)$ is the marginal distribution of X_1 and $f_k(x_k \mid x_1, \ldots x_{k-1})$ is the conditional density function of X_k with the condition $X_1 = x_1, \ldots, X_{k-1} = x_{k-1}$
- The idea is to generate the variables one at a time: first one draws value for X_1 from its marginal distribution, them one draws value for X_2 from the conditional distribution using the value of X_1 (already drawn) as the condition, etc.
- Denote by F_k the conditional cdf corresponding to the conditional pdf f_k and use the inverse transform method
- Algorithm
  - generate the random variables $U_1, \ldots, U_n$ from the uniform distribution $\mathrm{U}(0, 1)$
  - solve the equations (invert the cdf's)

$$
\begin{aligned}
F_1(X_1) &= U_1 \\
F_2(X_2 \mid X_1) &= U_2 \\
&\vdots \\
F_n(X_n \mid X_1, \ldots, X_{n-1}) &= U_n
\end{aligned}
$$

Multi-dimensional distribution: example

- The problem is to generate points (X, Y) in the unit square, with the left lower corner at the origin, using the density function which grows along the diagonal (the integral of the density over the square is 1)

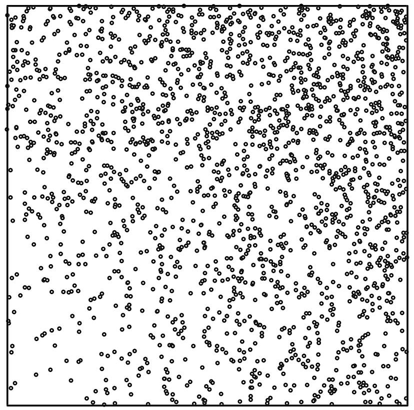
$$f(x, y) = x + y$$

- The marginal pdf and cdf of X are

$$f(x) = \int_0^1 f(x, y)\, dy = x + \frac{1}{2}, \qquad F(x) = \int_0^x f(x')\, dx' = \frac{1}{2}(x^2 + x)$$

- The conditional pdf and cdf functions of Y are

$$f(y \mid x) = \frac{f(x, y)}{f(x)} = \frac{x + y}{x + \frac{1}{2}}, \qquad F(y \mid x) = \int_0^y f(y' \mid x)\, dy' = \frac{xy + \frac{1}{2}y^2}{x + \frac{1}{2}}$$



- Inversion of the cdf functions gives the formulas

$$\begin{aligned} X &= \tfrac{1}{2}(\sqrt{8U_1 + 1} - 1) \\ Y &= \sqrt{X^2 + U_2(1 + 2X)} - X \end{aligned}$$

Generation from a multinormal distribution

- A random vector $\mathbf{X} = (X_1, \ldots, X_n)$, which obeys multi-dimensional normal distribution (multinormal distribution) has the pdf

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})}$$

  where $\mathbf{m}$ is the mean (vector) and $\mathbf{\Sigma}$ is the covariance matrix

- Since $\mathbf{\Sigma}$ is a positive definite and symmetric matrix one can always find a unique lower triangular matrix (alternatively a symmetric matrix) $\mathbf{C}$ such that $\mathbf{\Sigma} = \mathbf{C}\mathbf{C}^T$

- Samples of $\mathbf{X}$ can now be generated as follows

- Algorithm

$$\mathbf{X} = \mathbf{C}\mathbf{Z} + \mathbf{m}$$

  where the components of the vector $\mathbf{Z} = (Z_1, \ldots, Z_n)$ are independent normally distributed random variables, $Z_i \sim \mathrm{N}(0, 1)$

  - the formula can be verified by making a change of variables in the density function, whereby the pdf of $\mathbf{Z}$ is obtained as $(2\pi)^{-n/2} e^{-\frac{1}{2}\mathbf{Z}^T\mathbf{Z}}$