Stochastic Simulation
Generation of random variables
Continuous sample space

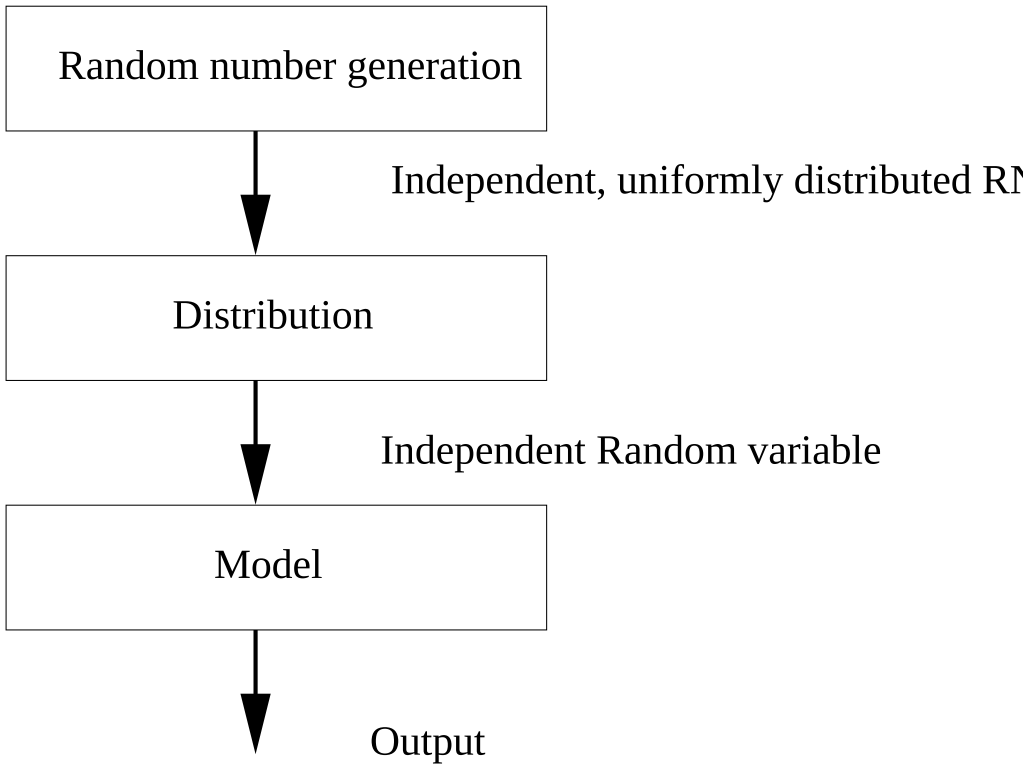
Institute of Mathematical Modelling

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.

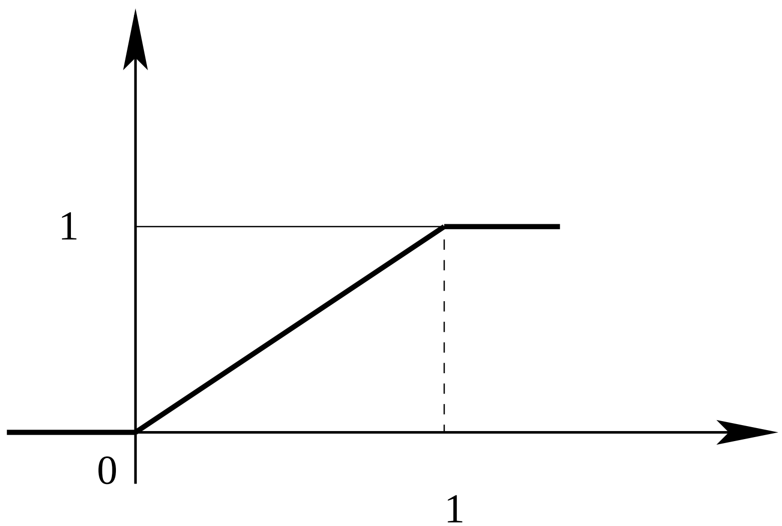
Generation of (pseudo)random variates

- Inverse transformation techniques
- Composition methods
- Acceptance/rejection methods
- Mathematical methods

Uniform distribution I

Our basic distribution or building block, $U_i \sim U(0,1)$

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

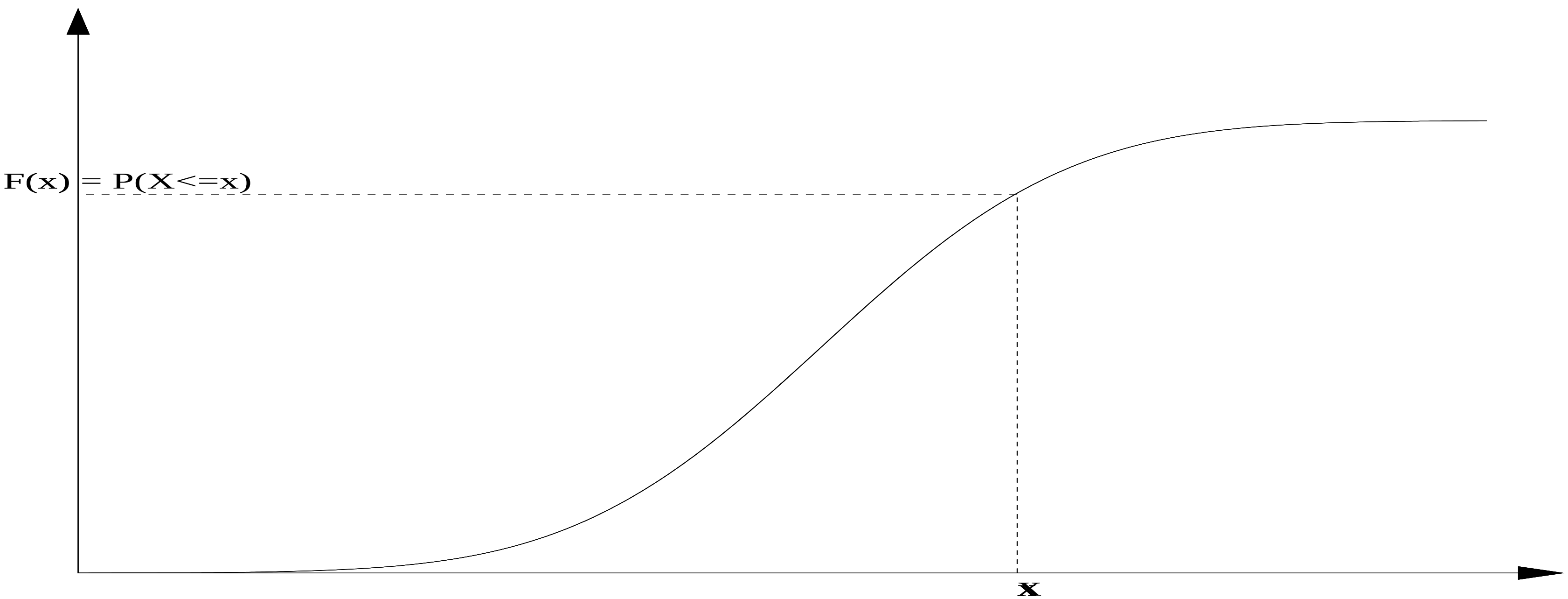


$$\mathsf{E}(U_i) = \frac{1}{2} \quad \mathsf{Var}(U_i) = \frac{1}{12}$$

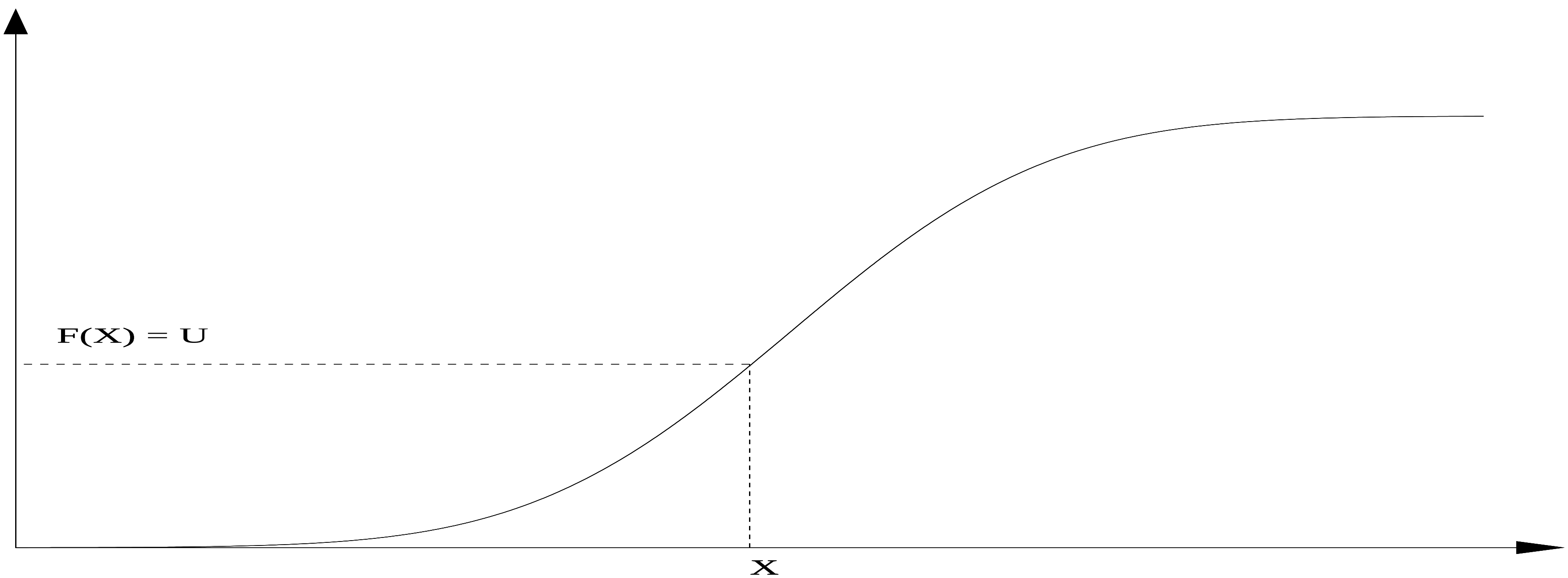
Inverse transformation

The cumulative distribution function (CDF)

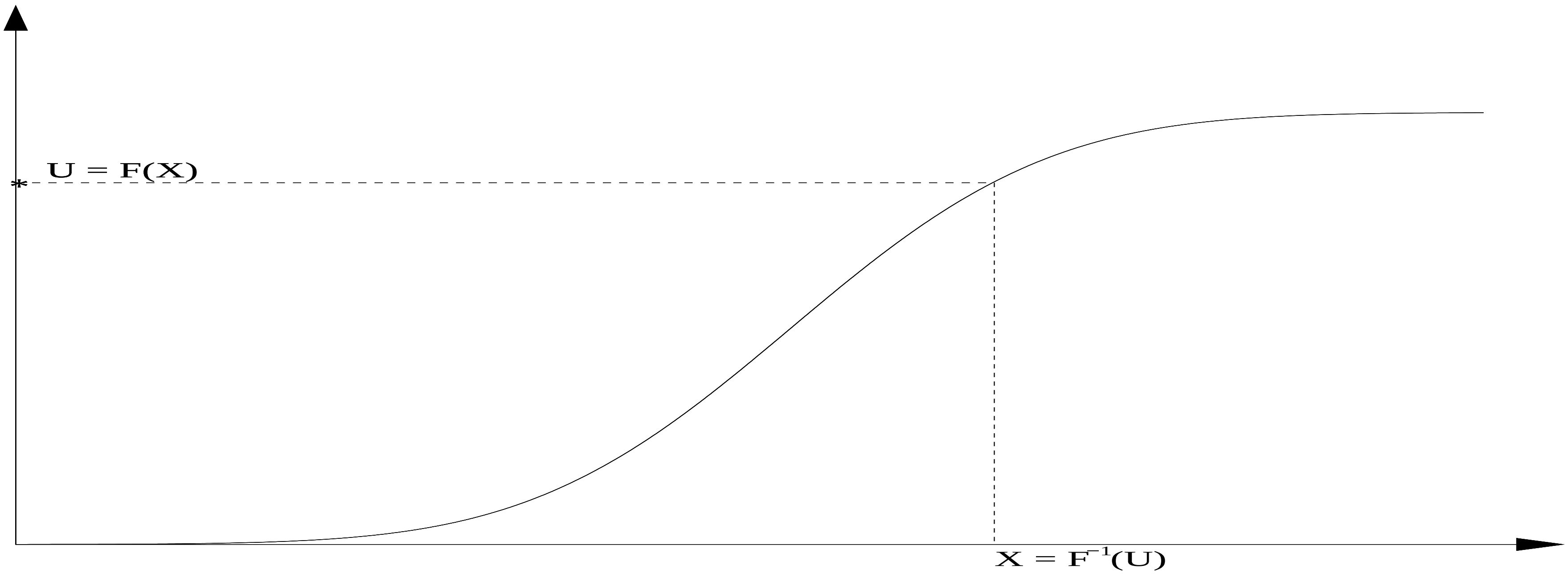
$$F(x) = \mathsf{P}(X \leq x)$$



The random variable $F(X)$

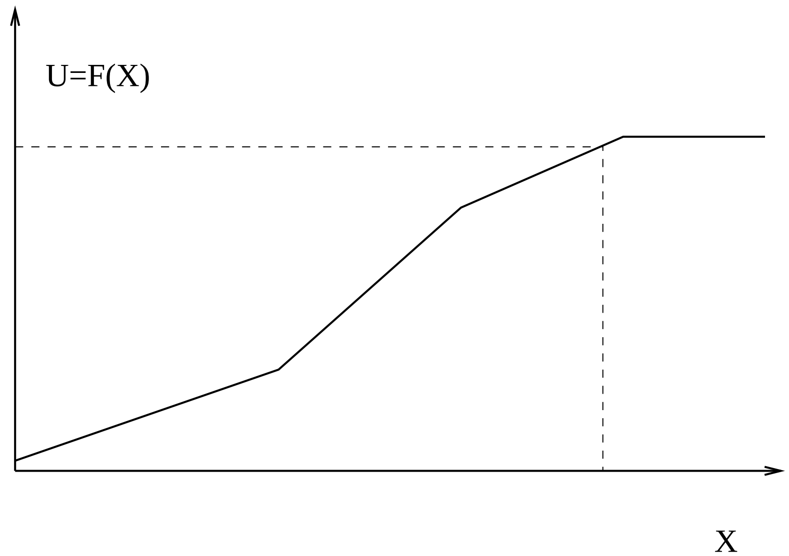


From U to X



Inversion method

The random variable $U = F(X)$



$$U = F(X) \qquad F(x) = \mathsf{P}(X \leq x)$$

$$\mathsf{P}(U \leq u) = \mathsf{P}(F(X) \leq u) = \mathsf{P}(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u$$

I.e. $F(X)$ is uniformly distributed.

So $1 - F(X)$ is also uniformly distributed - the p-value

Inversion method

Assume g continuous and increasing and let:

$$X = g(Y) \qquad Y \sim F_Y(y) = \mathsf{P}(Y \leq y)$$

then

$$F_X(x) = \mathsf{P}(X \leq x) = \mathsf{P}(g(Y) \leq x) = \mathsf{P}(Y \leq g^{-1}(x)) = F_Y(g^{-1}(x))$$

If $Y = U$ then $F_U(u) = u$, and $F_X(x) = g^{-1}(x)$.
If

$$\boxed{X = F^{-1}(U)}$$

then X will have the cdf $F(x)$ ($(F^{-1}(x))^{-1} = F(x)$).

Uniform distribution II

Now, focus on $U(a,b)$.

$$f(x) = \frac{1}{b-a} \qquad a \leq x \leq b$$

$$F(x) = \frac{x-a}{b-a} \qquad F^{-1}(u) = a + (b-a)u$$

$$X = a + (b-a)U \qquad \sim \qquad U(a,b)$$

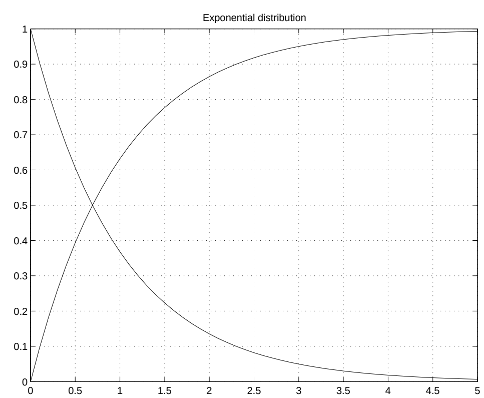
Exponential distribution

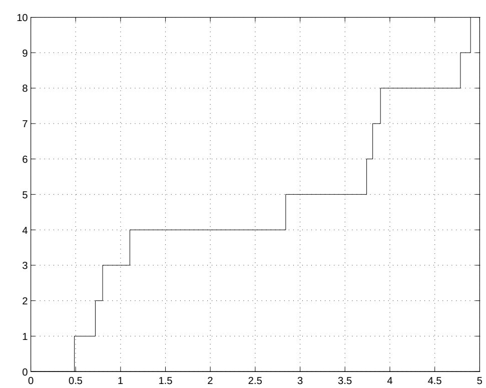
The time between events in a Poisson process is exponentially distributed. (Arrival time)

$$F(x) = 1 - e^{(-\lambda x)} \quad \mathsf{E}(X) = \frac{1}{\lambda} \quad F^{-1}(u) = -\frac{1}{\lambda}\log(1-u)$$

So (both $1-U$ and U are uniformly distributed)

$$X = -\frac{\log(U)}{\lambda} \sim \exp(\lambda)$$



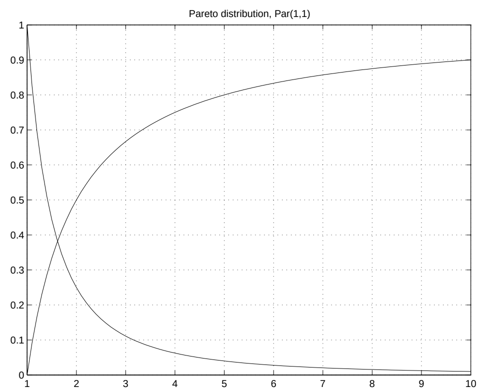


Pareto

Is often used in connection to description of income (over a certain level).

$$X \sim Pa(k,\beta) \quad F(x) = 1 - \left(\frac{\beta}{x}\right)^k \quad x \geq \beta \;\; X = \beta\left(U^{-\frac{1}{k}}\right)$$

$$\mathsf{E}(X) = \frac{k}{k-1}\beta \quad \mathsf{Var}(X) = \frac{k}{(k-1)^2(k-2)}\beta^2 \quad k > 1,\;\; 2$$



Pareto with $X \geq 0$

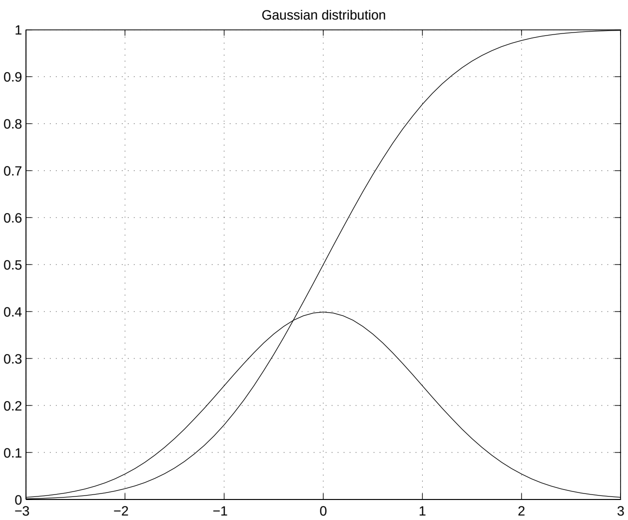
$$F(x) = 1 - \left(1 + \frac{x}{\beta}\right)^{-k} \quad X = \beta\left(U^{-\frac{1}{k}} - 1\right)$$

Gaussian

X a result of many (∞) independent sources (Central limit theorem)

$$X \sim \mathsf{N}(\mu, \sigma^2)$$

$$Z \sim \mathsf{N}(0,1) \qquad X = \mu + \sigma Z \qquad Z = \Phi^{-1}(U)$$



Rayleigh Fordeling

- For $X_i \sim \mathsf{N}(0,1)$, define $R = \sqrt{X_1^2 + X_2^2}$. Now

  $$F_R(r) = \mathsf{P}(R \leq r) = 1 - e^{-\frac{r^2}{2}}$$

  With $S = R^2 = X_1^2 + X_2^2$

  $$F_S(s) = \mathsf{P}(S \leq s) = 1 - e^{-\frac{s}{2}}$$

  i.e. $S \sim \exp\left(\frac{1}{2}\right)$.

- $X_1 = R\cos(2\pi\theta)$ and $X_2 = R\sin(2\pi\Theta)$ for $\Theta \sim \mathsf{Unif}(0;1)$ indepedent of R.

Mathematical Method

- By means of transformation and other techniques we can obtain a random variable with a certain distribution.

The Box-Muller method A transformation from polar ($\theta = 2\pi U_2$, $r = \sqrt{-2\log(U_1)}$) into Cartesian coordinates ($X = Z_1$ and $Y = Z_2$).

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \sqrt{-2\log(U_1)} \begin{bmatrix} \cos(2\pi U_2) \\ \sin(2\pi U_2) \end{bmatrix} \qquad Z_1,\ Z_2 \sim \mathsf{N}(0,1)$$

Central limit theorem

$$X = \sum_{i=1}^{n} U_i - \frac{n}{2} \qquad \text{eg.} \quad n = 6$$

Generation of $\cos$ and $\sin$

Sine and cosine can be calculated by the following acceptance/rejection algorith m:

1. Generate $V_1, V_2 \sim U(-1, 1)$
2. Generate $R^2 = V_1^2 + V_2^2$
3. If $R^2 > 1$ goto 1.
4. $\cos(2\pi U_2) = \frac{V_1}{R}, \sin(2\pi U_2) = \frac{V_2}{R}$

Logarithmic Gaussian, LN(α, β^2)

$$Y \sim \mathsf{LN}(\alpha, \beta^2) \qquad \log(Y) \sim \mathsf{N}(\alpha, \beta^2)$$

$$Y = e^X \qquad X = \alpha + \beta\ Z \qquad Z \sim \mathsf{N}(0, 1)$$

General and mulitvariate normal distribution

- Generate n independent values from an N$(0, 1)$ distribution,$Z_i \sim$ N$(0, 1)$.
- $X_i = \mu_i + \sum_{j=1}^{i} c_{ij} Z_j$
- Where c_{ij} are the elements in the Cholesky factorisation of $\Sigma, \Sigma = CC'$

Composition methods - hyperexponential distribution

$$F(x) = 1 - \sum_{i=1}^{m} p_i e^{-\lambda_i x} = \sum_{i=1}^{m} p_i \left(1 - e^{-\lambda_i x}\right)$$

Formally we can express

$Z = X_I$ where $I \sim \{1, 2, \ldots, m\}$ with $\mathsf{P}(I = i) = p_i$ and $X_I \sim \exp(\lambda_I)$

1. Choose $I \sim \{1, 2, \ldots, m\}$ with probabilities p_i's
2. $Z = -\frac{1}{\lambda_I} \log(U)$

Composition methods - Erlang distribution

- The Erlang distribution is a special case of the Gamma distribution with integer valued shape parameter
- An Erlang distributed random variable can be interpreted as a sum of independent exponential variables
- We can generate an Erlang-n distributed random variate by adding n exponential random variates.

$$Y \sim \mathsf{Erl}_n(\lambda) \qquad \mathsf{E}(Y) = \frac{n}{\lambda} \qquad \mathsf{Var}(Y) = \frac{n}{\lambda^2}$$

with $\lambda_i = \lambda$

$$Y = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} -\frac{1}{\lambda} \log\left(U_i\right) = -\frac{1}{\lambda} \log\left(\Pi_{i=1}^{n} U_i\right)$$

Composition methods II

Generalization:

$$f(x) = \int f(x|y) f(y) dy$$

$$X \text{ given } Y : f(x|y) \quad Y : f(y)$$

Y is typically a parameter (eg. the conditional distribution of X given $Y = \mu$ is $\mathsf{N}(\mu, \sigma^2)$)

Generate:

- Generate Y from $f(y)$.
- Generate X from $f(x|y)$ where Y is used.

Composition methods example of generalisation

$$f(y) = \mu e^{-\mu y}, \quad f_X(x|Y=y) = y e^{-yx}$$

$$\begin{aligned} f(x) &= \int_0^\infty y e^{-yx} \mu e^{-\mu y} \mathsf{d}y = \mu \int_0^\infty y e^{-(\mu+x)y} \mathsf{d}y = \frac{\mu}{(\mu+x)^2} \\ &= \frac{\frac{1}{\mu}}{\left(1+\frac{x}{\mu}\right)^2} \\ F(x) &= 1 - \left(1+\frac{x}{\mu}\right)^{-1} \end{aligned}$$

a Pareto distribution. The example can easily be generalised with gamma distributions, the algebra is slighty more involved.

Acceptance/rejection

Problem: we would like to generate X from pdf f, but it is much faster to generate Y
with pdf g. NB. X and Y have the same sample space. If

$$\frac{f(y)}{g(y)} \leq c \qquad \text{for all } y \text{ and some } c$$

- Step 1. Generate Y having density g.
- Step 2. Generate a random number U
- If $U \leq \frac{f(Y)}{cg(Y)}$ set $X = Y$.Otherwise return to step 1.

$$g(y)dy\frac{f(y)}{cg(y)} = \frac{f(y)dy}{c}$$

```
Statistics Toolbox
Version 4.0 (R13) 20-Jun-2002
     .
     .
     .
 Random Number Generators.
  betarnd     - Beta random numbers.
  binornd     - Binomial random numbers.
  chi2rnd     - Chi square random numbers.
  exprnd      - Exponential random numbers.
  frnd        - F random numbers.
  gamrnd      - Gamma random numbers.
  geornd      - Geometric random numbers.
  hygernd     - Hypergeometric random numbers.
  iwishrnd    - Inverse Wishart random matrix.
  lognrnd     - Lognormal random numbers.
```

Exercise 3

1. Generate simulated values from the following distributions
   (a) Exponential distribution
   (b) Normal distribution (at least with standard Box-Mueller)
   (c) Pareto distribution, with $\beta = 1$ and experiment with different values of k values: $k = 2.05$, $k = 2.5$, $k = 3$ and $k = 4$.

   Verify the results by comparing histograms with analytical results and perform tests for distribution type.
2. For the Pareto distribution with support on $[\beta, \infty[$ compare mean value and variance, with analytical results, which can be calculated as $\mathrm{E}(X) = \beta\frac{k}{k-1}$ (for $k > 1$) and $\mathrm{Var}(X) = \beta^2\frac{k}{(k-1)^2(k-2)}$ (for $k > 2$). Explain problems if any.
3. For the normal distribution generate 100 95% confidence intervals for the mean and variance, each based on 10 observations. Discuss the results.
4. Simulate from the Pareto distribution using composition.