Skoltech

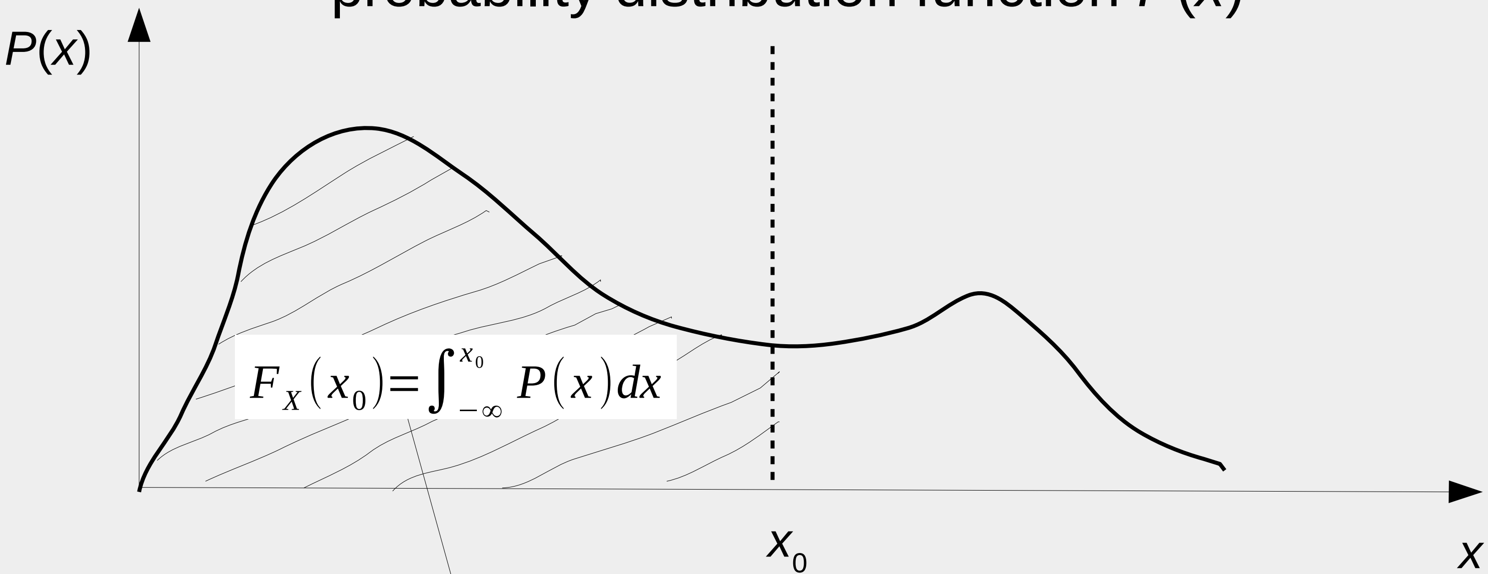
Skolkovo Institute of Science and Technology

Stochastic methods in Mathematical Modelling

Lecture 3. Distribution sampling 1

Distribution sampling

Cumulative distribution function for probability distribution function $P(x)$



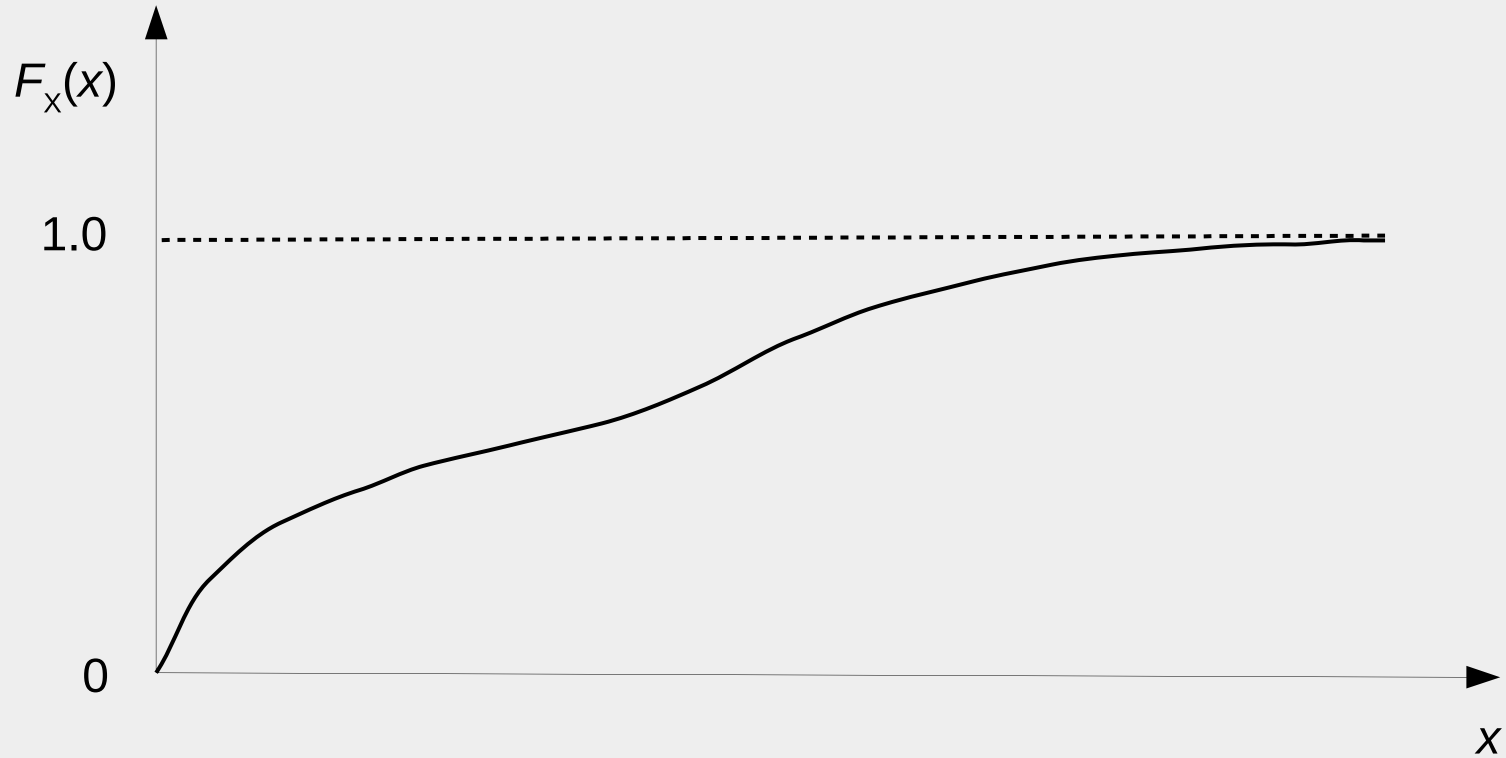
$$F_X(x_0) = \int_{-\infty}^{x_0} P(x)\,dx$$

Cumulative distribution function (CDF)

N.B. CDF can also be defined for discrete and mixed distributions

Distribution sampling

Cumulative distribution function $F_X(x)$



$$P(x) = \frac{dF_X(x)}{dx}$$ if the derivative exists

Skoltech
Skolkovo Institute of Science and Technology

Distribution sampling

How does one sample random variables from a distribution by using CDF $F_x(X)$?

Let's define $$Y = F_X(X)$$

Notice that $F_X(x)$ transforms a function defined on *R* to the function defined on [0,1]. $F_X^{-1}(x)$ is unique and exists since CDF is monotonous. Then

$$F_Y(y) = \Pr(Y \le y) = \Pr(X \le F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y, y \in [0,1]$$

$$f_Y(y) = \begin{cases} 1 & y \in [0,1] \\ 0 & \text{otherwise.} \end{cases}$$

Y is uniform distribution on [0,1]

$$Y \sim U([0,1])$$

Skoltech
Skolkovo Institute of Science and Technology

Distribution sampling

How does one sample random variables from a distribution by using CDF?

Inverse transform sampling. The algorithm

1. Find the inverse function of the CDF, $F_X^{-1}(x)$

2. Generate a random number *u* from U[0,1]

3. Compute $X = F_X^{-1}(u)$. The computed random variable X has the sought distribution

Distribution sampling

How does one sample random variables from a distribution by using CDF?

The example. The exponential PDF $P(x)=\lambda e^{-\lambda x}$

The CDF for exponential PDF reads $F_X(x)=1-e^{-\lambda x}$

$$x = F_X^{-1}(y) = \frac{-1}{\lambda}\ln(1-y)$$

Since y belongs to [0,1] one can simply generate the exponential with

$$x = F_X^{-1}(y) = \frac{-1}{\lambda}\ln(y)$$

y is drawn from U[0,1]

Q: Do this for $P(x) \sim 1/(1+x^2)$, $-\infty \le x < \infty$

Distribution sampling

Q. Find a way to sample the values from a distribution with the PDF $P(x) \sim 1/(1+x^2)$, $-\infty \leq x < \infty$

Cauchy distr. sampling

CDF: $F_X(x) = \int_{-\infty}^{x} \frac{dx}{\pi(1+x^2)} = \frac{1}{\pi} \arctan x + \frac{1}{2} = y$

$x = \tan\left(\pi\left(y - \frac{1}{2}\right)\right)$

Distribution sampling

Box-Muller transform for generation of Gaussian distributions

For the Gaussian case it would be difficult to use the method above due to the lack of analytic form for $F_X^{-1}(x)$

Let's take U_1 and U_2 are uniformly distributed RVs on [0,1]

Then Z_0 and Z_1 are two independently distributed Gaussian RVs from *N*[0,1]

$$Z_0=\sqrt{-2\ln U_1}\sin(2\pi U_2)$$

$$Z_1=\sqrt{-2\ln U_1}\cos(2\pi U_2)$$

Skoltech
Skolkovo Institute of Science and Technology

Distribution sampling

Joint probability distribution of functions of random variables

1. X_1 and X_2 are continuous RVs with joint PDF $f_{X1,X2}(x_1,x_2)$
2. $Y_1=g_1(X_1,X_2)$, $Y_2=g_2(X_1,X_2)$ and there is a *unique* solution $x_1=h_1(y_1,y_2)$, $x_2=h_2(y_1,y_2)$
3. Also g_1 and g_2 have continuous PDFs at all points x_1,x_2 i.e.

$$J(X_1, X_2) = \begin{vmatrix} \dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} \\ \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{\partial x_2} \end{vmatrix} \neq 0$$

then $$\boxed{f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)\,|J(x_1, x_2)|^{-1}}$$

Distribution sampling

Joint probability distribution of functions of random variables

Joint PDFs of RVs

RVs: $X_1, X_2 \longrightarrow Y_1, Y_2$

$$Y_1 = g_1(X_1, X_2)$$
$$Y_2 = g_2(X_1, X_2)$$

assumptions:

1) $y_1 = g_1(x_1, x_2)$, $y_2 = g_2(x_1, x_2)$ $\longrightarrow$ uniques solution $x_1 = h_1(y_1, y_2)$, $x_2 = h_2(y_1, y_2)$

2) g_1, g_2 have continuous partial derivatives for all x_1, x_2 and

$$J_{x_1, x_2} = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} \neq 0$$

then

$$f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(x_1, x_2) \left| J(x_1, x_2) \right|^{-1}$$

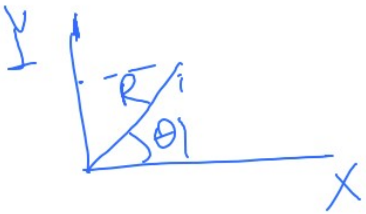
$$P\{Y_1 \le y_1, Y_2 \le y_2\} = \iint\limits_{\substack{x_1, x_2: \\ g_1(x_1, x_2) \le y_1 \\ g_2(x_1, x_2) \le y_2}} f_{X_1 X_2}(x_1, x_2)\, dx_1\, dx_2$$

Skoltech
Skolkovo Institute of Science and Technology

Distribution sampling

Box-Muller transform derivation

Box-Muller transform



$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

$$d = g_1(x,y) = x^2 + y^2 = R^2$$

$$\theta = g_2(x,y) = \arctan(y/x)$$

$$J = \begin{vmatrix} 2x & 2y \\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{vmatrix} = 2$$

$$f(d,\theta) = \frac{f(x,y)}{2} = \underbrace{\frac{1}{2} e^{-\frac{d}{2}}}\ \underbrace{\frac{1}{2\pi}}, \quad 0 < d < \infty, \quad 0 \le \theta \le 2\pi$$

$f(d)f(\theta) \rightarrow R^2$ & θ are independent U_1, U_2 are $U[0,1]$

$R^2 = -2 \ln U_1$ since $X = R\cos\theta$, $Y = R\sin\theta$ $\rightarrow \begin{cases} X = \sqrt{-2\ln U_1}\cos(2\pi U_2) \\ Y = \sqrt{-2\ln U_1}\sin(2\pi U_2) \end{cases}$

$\theta = 2\pi U_2$