RANDOM VARIABLE

RANDOM VARIABLE ...

RANDOM VARIABLE — A 'MAPPING' OF THE SET OF (ELEMENTARY) EVENTS E onto the set of real numbers $\mathcal{R}$.

For instance:

- height of a person met in the street;
- number of people in Cracow down with flu each day;
- number of meteorites falling each year per 1 km^2;
- number of minutes you wait every day for the street-car;
- number of accidents per months at a given street-intersection;
- strength of a climbing-rope;
- number of deaths in Cracow in (each) November
- **a result of every measurement**.

RANDOM VARIABLE . . .

RANDOM VARIABLE . . .

and its (CUMULATIVE) DISTRIBUTION FUNCTION

RANDOM VARIABLE — A "MAPPING" OF THE SET OF (ELEMENTARY) EVENTS E onto the set of real numbers $\mathcal{R}$

RANDOM VARIABLE . . .

and its (CUMULATIVE) DISTRIBUTION FUNCTION

RANDOM VARIABLE — A "MAPPING" OF THE SET OF (ELEMENTARY) EVENTS E onto the set of real numbers $\mathcal{R}$

$X \equiv$ RANDOM VARIABLE; x — ITS VALUE (REALISATION)

RANDOM VARIABLE . . .

and its (CUMULATIVE) DISTRIBUTION FUNCTION

RANDOM VARIABLE — A "MAPPING" OF THE SET OF (ELEMENTARY) EVENTS E onto the set of real numbers $\mathcal{R}$

$X \equiv$ RANDOM VARIABLE; x — ITS VALUE (REALISATION)

we introduce cumulative distribution function: $F_X(x)$ (or shortly: $F(x)$) as

$$\boxed{F_X(x) \equiv F(x) = \mathcal{P}(X \leq x)}$$

Some textbooks use a slightly different definition

$$\boxed{F_X(x) \equiv F(x) = \mathcal{P}(X < x)}$$

It has no any influence in the case of continuous RV; but for a discrete RV **it makes quite a difference**

DISTRIBUTION FUNCTION $F(x)$

and its properties;

1. $0 \leq F(x) \leq 1$ FOR EVERY x;

DISTRIBUTION FUNCTION $F(x)$

and its properties;

1. $0 \leq F(x) \leq 1$ FOR EVERY x;
2. $\lim_{x\to-\infty} \equiv F(-\infty) = 0 \quad \lim_{x\to\infty} \equiv F(\infty) = 1$;

DISTRIBUTION FUNCTION $F(x)$

and its properties;

1. $0 \leq F(x) \leq 1$ FOR EVERY x;
2. $\lim_{x \to -\infty} \equiv F(-\infty) = 0 \quad \lim_{x \to \infty} \equiv F(\infty) = 1$;
3. $F(x)$ IS A NON-DECREASING FUNCTION;

DISTRIBUTION FUNCTION $F(x)$

and its properties;

1. $0 \leq F(x) \leq 1$ FOR EVERY x;
2. $\lim_{x\to-\infty} \equiv F(-\infty) = 0 \quad \lim_{x\to\infty} \equiv F(\infty) = 1$;
3. $F(x)$ IS A NON-DECREASING FUNCTION;
4. $F(x)$ IS RIGHT-SIDED (AT LEAST) CONTINUOUS: $F(x+0) = F(x)$;

DISTRIBUTION FUNCTION $F(x)$

and its properties;

1. $0 \leq F(x) \leq 1$ FOR EVERY x;
2. $\lim_{x \to -\infty} \equiv F(-\infty) = 0 \quad \lim_{x \to \infty} \equiv F(\infty) = 1$;
3. $F(x)$ IS A NON-DECREASING FUNCTION;
4. $F(x)$ IS RIGHT-SIDED (AT LEAST) CONTINUOUS: $F(x+0) = F(x)$;
5. $\mathcal{P}(a < X \leq b) = F(b) - F(a)$;

DISTRIBUTION FUNCTION $F(x)$

and its properties;

1. $0 \le F(x) \le 1$ FOR EVERY x;
2. $\lim_{x\to-\infty} \equiv F(-\infty) = 0 \quad \lim_{x\to\infty} \equiv F(\infty) = 1$;
3. $F(x)$ IS A NON-DECREASING FUNCTION;
4. $F(x)$ IS RIGHT-SIDED (AT LEAST) CONTINUOUS: $F(x+0) = F(x)$;
5. $\mathcal{P}(a < X \le b) = F(b) - F(a)$;
6. $\mathcal{P}(X = x_0) = F(x_0) - F(x_0 - 0)$

RANDOM VARIABLE OF THE DISCRETE TYPE . . .

. . . and the probability distribution function:

RANDOM VARIABLE OF THE DISCRETE TYPE ...

...and the probability distribution function:

there exists a finite (or at least enumerable) set of the values $\{x_1, \ldots, x_k, \ldots\}$ of the random variable X
which occur with probabilities $\{p_1, \ldots, p_k, \ldots\}$:

RANDOM VARIABLE OF THE DISCRETE TYPE . . .

. . . and the probability distribution function:

there exists a finite (or at least enumerable) set of the values $\{x_1, \ldots, x_k, \ldots\}$ of the random variable X
which occur with probabilities $\{p_1, \ldots, p_k, \ldots\}$:

$$\mathcal{P}(X = x_i) = p_i \; > \; 0; \quad \sum_{i=1} p_i = 1$$

The set of all $p(x_i)$ values is called *probability distribution* or *probability function*

...and the probability distribution function:

there exists a finite (or at least enumerable) set of the values $\{x_1, \ldots, x_k, \ldots\}$ of the random variable X which occur with probabilities $\{p_1, \ldots, p_k, \ldots\}$:

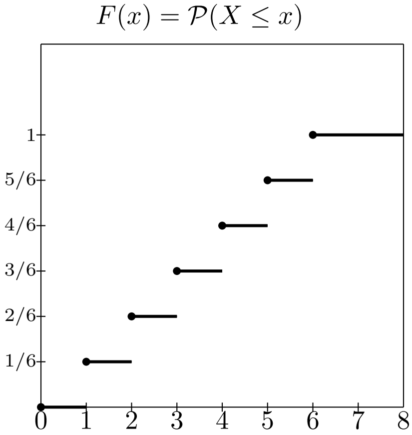
$$\mathcal{P}(X = x_i) = p_i > 0; \quad \sum_{i=1} p_i = 1$$

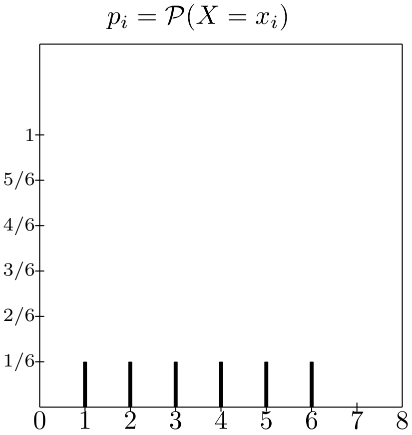
The set of all $p(x_i)$ values is called *probability distribution* or *probability function*

Our cumulative (probability) distribution is given as:

$$F(x) = \mathcal{P}(X \leq x) = \sum_{-\infty < x_i \leq x} p_i$$

cumulative distribution function (left) and probability function (right)





RANDOM VARIABLE OF THE CONTINUOUS TYPE

There exists : $f(x)$ for $-\infty < x < \infty$; $\quad f(x) \geq 0$, which is related to $F(x)$ as

$$F(x) = \int_{-\infty}^{x} f(s)\, ds$$

RANDOM VARIABLE OF THE CONTINUOUS TYPE

There exists : $f(x)$ for $-\infty < x < \infty$; $f(x) \geq 0$, which is related to $F(x)$ as

$$F(x) = \int_{-\infty}^{x} f(s)\, ds$$

The two functions have the following properties:

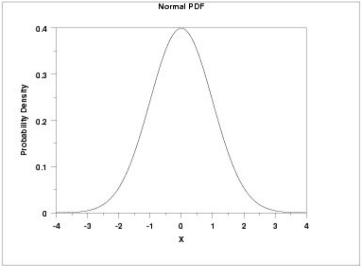
1. $\frac{dF}{dx} = f(x) \qquad F(x) = \int_{-\infty}^{x} f(s)\, ds;$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1;$
3. $\bigwedge_{c \in R} \mathcal{P}(X = c) = 0;$
4. $$\mathcal{P}(a \leq X < b) = \mathcal{P}(a < X \leq b) = \mathcal{P}(a < X < b) = \mathcal{P}(a \leq X \leq b)$$
$$= F(b) - F(a) = \int_{a}^{b} f(x)\, dx;$$

We call $f(x)$ — the probability density function

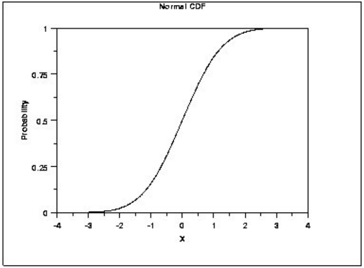
$$\boxed{\mathcal{P}(X \in (x, x + dx)) = f(x)dx.}$$

RANDOM VARIABLE and NORMAL DISTRIBUTION

here come graphs of the pdf AND cdf for the standardised normal distribution:



http://www.itl.nist.gov; Jan 5th 2012



http://www.itl.nist.gov; Jan 5th 2012

- play: Wolfram's BELL CURVES
- play: Wolfram's Standard Normal Distribution Areas

CHANGE OF VARIABLE

Suppose we ave a RV X of a continuous type and we know its pdf $f(x)$. Now, we have another RV that is functionally related to X:

$$Y = Y(X).$$

Can we say anything about the pdf for Y, $g(y)$?

CHANGE OF VARIABLE

Suppose we ave a RV X of a continuous type and we know its pdf $f(x)$. Now, we have another RV that is functionally related to X:

$$Y = Y(X).$$

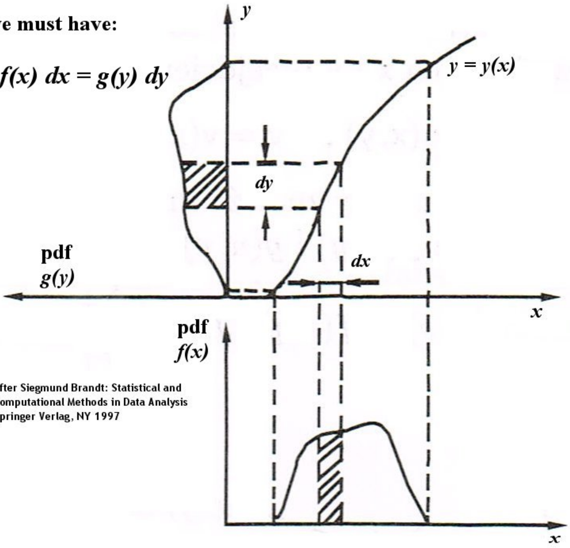
Can we say anything about the pdf for Y, $g(y)$?
Simpler case: Y is a monotonic function of X. Then, from a simple geometrical reasoning (cf. the picture – next slide):

$$g(y) = \left|\frac{dx}{dy}\right| f(x)$$

The dx/dy is the derivative of X with respect to Y. Of course we have $\frac{dy}{dx} = \left(\frac{dx}{dy}\right)^{-1}$. For a non-monotonic $y = y(x)$ dependence one must take into account that different regions of the X variable may be mapped into one (the same) region of the Y variable. The $g(y)$ pdf in such a region will be a sum of $f(x)$ pdf's multiplied by $|dx/dy|$ over all the regions of X which have been mapped into the given region of Y.
We shall return to this question when we will be more acquainted with some *types of distributions of Rvs*.

we must have:

$$f(x)\, dx = g(y)\, dy$$



after Siegmund Brandt: Statistical and Computational Methods in Data Analysis Springer Verlag, NY 1997