MECH 534 Computer-Based Modeling and Simulation
Notes on *Probability Density and*
Cumulative Distribution Functions
Prof. Cagatay Basdogan

Introduction

Probability concepts are familiar parts of everyday life. Roughly, the probability of an event is our *degree of belief* in that event actually occurring. Simple properties of probabilities are:

- probabilities are numbers between 0 and 1;
- the closer the probability is to 0, the more unlikely is the associated event to occur;
- the closer the probability is to 1, the more sure we are that the associated event will happen;
- the probability of an event is the sum of the probabilities of all the distinct outcomes making up the event; and
- the probabilities of all distinct possible outcomes sum to one.

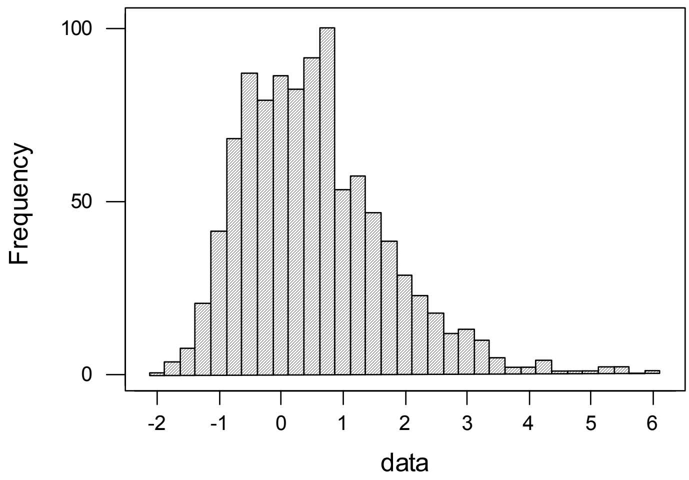
Thus an *impossible* event has probability is zero, and a *certain* event has probability one.

1. The probability density function (pdf)

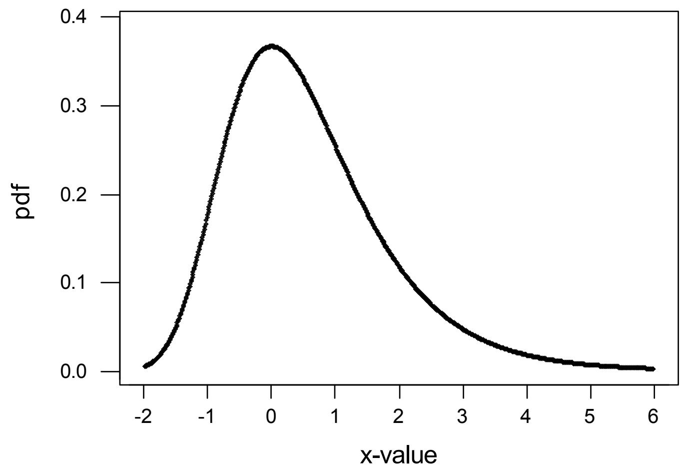
The theoretical, or underlying form of a distribution of data can be described by a *probability density function*, or *pdf*. In general, a *pdf* is the limiting form of histograms, when the number of observation grows infinitely large, and the class intervals become correspondingly small, so that a smooth curve results. The interpretation of histograms and *pdf* curves for continuous variables is that, when the horizontal and vertical scales are adjusted so that the total area under the curve is one, then probabilities are areas between the curve and the x-axis.

To illustrate, the histogram below is based upon 1000 observations from a certain distribution. The tendency to approach a smooth curve is clear, and the ultimate limiting form of the curve, which is the *pdf*, is shown on the next graph.

Histogram of 1000 observations



pdf as the limit of histograms

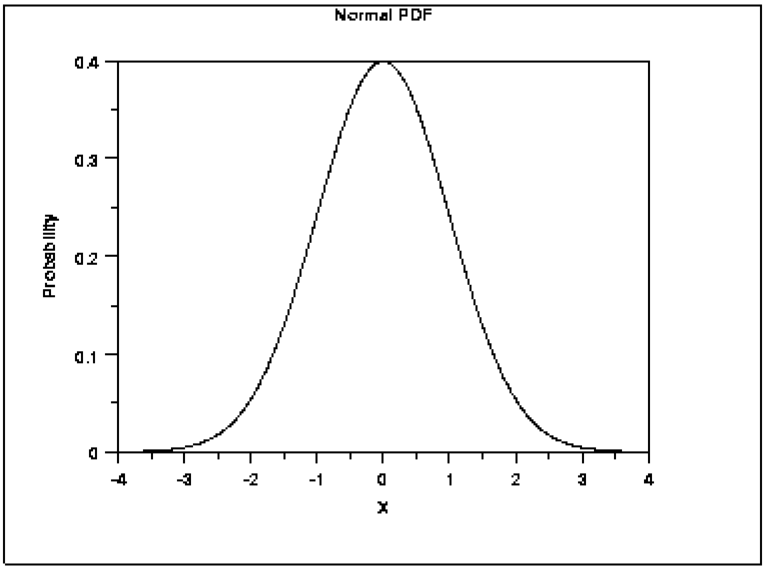


For a continuous function, the probability density function (pdf) is the probability that the variate has the value x. Since for continuous distributions the probability at a single point is zero, this is often expressed in terms of an integral between two points.

$$\int_a^b f(x)dx = Pr[a \le X \le b]$$

For a discrete distribution, the pdf is the probability that the variate takes the value x.

$$f(x) = Pr[X = x]$$



The plot of the normal density function.

2. The cumulative distribution function (cdf)

The cumulative distribution function (cdf) is the probability that the variable takes a value less than or equal to x. That is

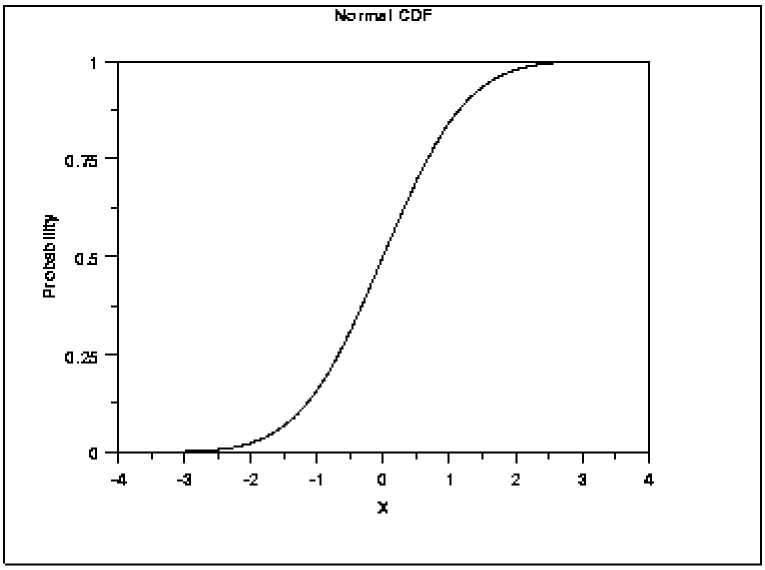
$$F(x) = Pr[X \leq x] = \alpha$$

For a continuous distribution, this can be expressed mathematically as

$$F(x) = \int_{-\infty}^{x} f(\mu)d\mu$$

For a discrete distribution, the cdf can be expressed as

$$F(x) = \sum_{i=0}^{x} f(i)$$



The plot of the normal cumulative distribution function.

3. Types of Probability Density Functions

1. **Gaussian (Normal) pdf.** - Use a "Gaussian-type" or normal distribution for most variables that have been scientifically measured or derived, or represent classifications of plants, animals or humans. Significant disadvantages to the Gaussian distribution function are that it is symmetric about its "mean" and spans the domain of, and is normalized over, all real numbers (from - infinity to + infinity). The following is a plot of the normal probability density function.

2. **Poisson pdf.** - Use a "Poisson-type" distribution for the failure of manufacturing or "process" components, or any rare phenomena. It serves as a useful probability model for events occurring randomly over time or space when all that is known is the average number of occurrences per unit time or space. The only input required is the average rate of occurrences per unit time or space. The "variance" is always equal to the "mean" of a Poisson distribution.

3. **Flat pdf.** - Use a constant value (a flat distribution curve) if there is thought to be little or no variation in the distribution.

4. **Triangular pdf.** - Use a triangular distribution if you know very little about the distribution. The Triangular Distribution Function can be completely defined by knowing the absolute minimum value, the most likely value, and the absolute maximum value.

5. **Other "Angular" pdfs** - Use an "angular" distribution that is derived to be a good fit to the data represented by the more conventional Distribution Functions (Gaussian, Poisson, etc.) These angular Distribution Functions are constructed by a series of straight-line approximations to the more complex continuous distribution functions.

6. **Reaction, or cross-section, pdfs** - These are used in physics and chemistry problems. They are tabulated for both chemical and nuclear reactions.